RISE OF THE ETHICAL MACHINES

BY

MATTHEW R. DOMSALLA

A THESIS SUBMITTED TO THE FACULTY OF
THE SCHOOL OF ADVANCED AIR AND SPACE STUDIES
FOR COMPLETION OF GRADUATION REQUIREMENTS

SCHOOL OF ADVANCED AIR AND SPACE STUDIES
AIR UNIVERSITY
MAXWELL AIR FORCE BASE, ALABAMA
JUNE 2012

APPROVAL

The undersigned certify that this thesis meets master's-level standards of research, argumentation, and expression.

_____

TIMOTHY P. SCHULTZ, Col, USAF, PhD (Date)

_____

MICHAEL W. KOMETER, Col, USAF, PhD (Date)

DISCLAIMER

The conclusions and opinions expressed in this document are those of the author. They do not reflect the official position of the US Government, Department of Defense, the United States Air Force, or Air University.

ABOUT THE AUTHOR

Prior to attending the School of Advanced Air and Space Studies, Lieutenant Colonel Matthew R. Domsalla was an Air Force Fellow at the Defense Advanced Research Projects Agency (DARPA) where he served as an Air Force advisor to the Persistent Close Air Support, Transformer, Triple Target Terminator, and Hypersonic Test Vehicle-2 programs. He graduated from the US Air Force Academy in 1997 with a Bachelor of Science degree in Mechanical Engineering and a minor in Mathematics. Following his commissioning, he attended Harvard University's John F. Kennedy School of Government, where he completed a Master's of Public Policy degree in International Security and Public Policy. He then attended Specialized Undergraduate Pilot Training at Laughlin Air Force Base, Texas. After completing pilot training and A-10 initial qualification training, he was assigned to the 75th Fighter Squadron at Pope Air Force Base, North Carolina. During this assignment, he flew combat sorties as part of Operations SOUTHERN WATCH, ENDURING FREEDOM, and IRAQI FREEDOM. He was then assigned to the 25th Fighter Squadron at Osan Air Base, South Korea.

In January 2007, Lieutenant Colonel Domsalla was selected to attend the US Air Force Test Pilot School at Edwards Air Force Base, California. Following graduation, he served as an A-10 and F-16 Experimental Test Pilot in the 40th Flight Test Squadron at Eglin Air Force Base, Florida. At Eglin, he was involved in numerous A-10, F-15, and F-16 test programs. He was responsible for the developmental test and evaluation of A-10C Software Suites 3, 5, and 6, the integration of the GBU-54 on the A-10C, the initial certification of alternative fuels on the A-10, and F-16 flight envelope expansion.

Lieutenant Colonel Domsalla is a Senior Pilot with 1,900 flying hours in the A-10A/C, F-16C/D, and more than 25 additional aircraft types. Following the School of Advanced Air and Space Studies, he will assume duties as the Director of Operations for the 452d Flight Test Squadron at Edwards Air Force Base, California.

ACKNOWLEDGEMENTS

I would like to thank Colonel Timothy Schultz and Colonel Michael Kometer for their insights, guidance, and feedback on this work. I would particularly like to thank my family for their love, patience, and understanding while I completed this endeavor.

ABSTRACT

This study examines the moral, ethical, and legal issues surrounding the development of autonomous systems capable of employing lethal force. It explores the international law principles that will inform questions concerning the legality of these systems along with the moral and ethical arguments both for and against these systems. It then assesses the implications of the current approaches for developing an ethical reasoning capability for a machine along with the necessity of establishing trust in such systems. Without the trust that autonomous systems will function as designed, operators will be reluctant to employ these systems. Finally, the work evaluates the implications of human interactions with autonomous systems, particularly the underlying, and possibly unintentional, moral and ethical consequences of design choices. It argues that while senior political and military leaders will make the final decision to employ these systems, those involved in the process must assess the moral and ethical consequences associated with the development of these systems rather than donning ethical blinders while privileging technical issues.

CONTENTS

Illustrations

Table

Figures

Introduction

Reliance on advanced technology has been a central pillar of the American way of war, at least since World War II.

—Thomas G. Mahnken
Technology and the American Way of War Since 1945

Advances in artificial intelligence have led to the development of systems with increasing levels of autonomy. In 2007, the Defense Advanced Research Projects Agency (DARPA) held the Urban Challenge. For this competition, teams from across the country developed autonomous vehicles capable of interacting with traffic in an urban environment.[1] Within three short years, Google announced it had developed a fleet of autonomous cars that had logged over 140,000 miles on California highways.[2] In February 2011, after its intelligent analytic system Watson defeated two human players in a game of *Jeopardy!*, IBM announced it was adapting Watson's technology to aid in data analysis and decision-making in the medical field.[3] Research into autonomous military systems is making similar strides.

While military systems have increased in their autonomous capabilities, the employment of lethal force has remained contingent upon a human decision to authorize weapons employment. For example, Samsung Techwin developed the optionally armed SGR-1 robot, shown in Figure 1, to provide security around sensitive sites.[4]

---

[1] DARPA, "Urban Challenge," http://archive.darpa.mil/grandchallenge/index.asp (accessed 26 March 2012).
[2] Tom Vanderbilt, "Let the Robot Drive: The Autonomous Car of the Future is Here," *Wired*, 20 January 2012, http://www.wired.com/magazine/2012/01/ff_autonomouscars/all/1 (accessed 26 March 2012).
[3] Peter Pachal, "IBM's Watson Wins Jeopardy! Next Up: Fixing Health Care," *PC Magazine*, 16 February 2011, http://www.pcmag.com/article2/0,2817,2380489,00.asp (accessed 26 March 2012).
[4] Samsung Techwin, "SGR Series," http://www.samsungtechwin.com/product/product_01_01.asp (accessed 26 March 2012).

Figure 1: Samsung Techwin SGR-1 Robot

Source: Samsung Techwin, "SGR Series," http://www.samsungtechwin.com/product/product_01_01.asp (accessed 26 March 2012)

The South Korean military is currently utilizing the SGR-1 to monitor the Korean Demilitarized Zone. While the SGR-1 autonomously detects intruders, a human operator at a remote monitoring station is currently required for the SGR-1 to fire its weapons.[5]

Technological advancements in artificial intelligence have created the potential for a system to employ lethal force autonomously–without a human actively consenting to the decision. This possibility has tremendous moral and ethical implications. In fact, the USAF *Unmanned Aircraft Systems Flight Plan 2009 – 2047* asserts,

> Authorizing a machine to make lethal combat decisions is contingent upon political and military leaders resolving legal

[5] Kim Deok-hyun, "Army Test Machine-Gun Sentry Robots in DMZ," *Yonhap News Agency*, 13 July 2010, http://english.yonhapnews.co.kr/national/2010/07/13/14/0301000000AEN20100713007800315F.HTML# (accessed 26 March 2012).

> and ethical questions. These include the appropriateness of machines having this ability, under what circumstances it should be employed, where responsibility for mistakes lies and what limitations should be placed upon the autonomy of such systems … Ethical discussions and policy decisions must take place in the near term in order to guide the development of future UAS [unmanned aircraft systems] capabilities, rather than allowing the development to take its own path apart from this guidance.[6]

This investigation will examine the legal and ethical implications of the approaches currently proposed for developing autonomous systems capable of employing lethal force.

Legal and ethical discussions about technological developments are critical. Peter Singer, Director of the 21st Century Defense Initiative at the Brooking Institute, argues, "While our understanding of law and ethics moves at a glacial place, technology moves at an exponential pace."[7] Echoing this sentiment, Harry Yarger, Professor of National Security Policy at the US Army War College, asserts, "Technology often outruns political and strategic maturity, creating strategic conditions or consequences that neither are prepared to deal with appropriately."[8] Despite the persistence of moral and ethical questions concerning military systems capable of employing lethal force autonomously, their development continues. Therefore, military and engineering professionals must understand the moral and ethical concerns related to these systems to inform policy-makers as they attempt to resolve and reconcile these concerns.

---

[6] Department of the Air Force, *United States Air Force Unmanned Aircraft Systems Flight Plan 2009 – 2047* (Washington, DC: Headquarters, United States Air Force, 18 May 2009), 41. The US Air Force now uses the term Remotely Piloted Aircraft (RPA) for tele-operated systems such as the MQ-1 Predator, MQ-9 Reaper, or RQ-4 Global Hawk. Depending on the source, analogous terms include uninhabited aerial vehicle, unmanned aerial vehicle, and more generally, unmanned systems and unmanned vehicles.

[7] Peter W. Singer, "The Ethics of Killer Applications: Why Is It So Hard To Talk About Morality When It Comes to New Military Technology?" *Journal of Military Ethics* 9, no 4 (December 2010), 308.

[8] Harry R. Yarger, *Strategy and the National Security Professional: Strategic Thinking and Strategy Formulation in the 21st Century* (Westport, CT: Praeger, 2008), 143.

The American Way of War

The development of a military system capable of conducting an autonomous lethal engagement reflects a central characteristic of how the United States conducts modern warfare. Prior to the conclusion of World War II, a "strategy of annihilation" characterized the American way of war.[9] Yet a central problem for American strategists has been securing "victory in its desired fullness without paying a cost so high that the cost would mock the very enterprise of waging war."[10] The development of nuclear weapons has only exacerbated this problem. Since World War II, the *modus operandi* of the American military, particularly the US Air Force, has been the pursuit of advanced technology. In his 1983 National Book Award winning work *National Defense*, James Fallows, *Atlantic* magazine Washington correspondent, argues, "The distinguishing feature of modern American defense has been the pursuit of the magic weapon."[11] This predilection is even more pronounced in the US Air Force, leading RAND military analyst Carl Builder to describe the Air Force as worshiping at "the altar of technology."[12]

These tendencies emerge from the American strategic culture. Colin Gray, the renowned British strategic thinker and Professor of International Relations and Strategic Studies at the University of Reading, identifies eight characteristics, listed in Table 1, of American strategic culture.[13]

---

[9] Russel F. Weigley, *The American Way of War: A History of United States Military Strategy and Policy* (Bloomington, IN: Indiana University, 1973), xxii.
[10] Weigley, *The American Way of War*, xxii.
[11] James Fallows, *National Defense* (New York: Random House, 1981), 35.
[12] Carl Builder, "Service Identities and Behavior," in *American Defense Policy*, eds. Peter L. Hays, Brenda J. Vallance, and Alan R. Van Tassel (Baltimore, MD: Johns Hopkins University, 1997), 112.
[13] Colin S. Gray, *Explorations in Strategy* (Westport, CT: Praeger, 1996), 89 – 98.

Table 1: Colin Gray's Characteristics of American Strategic Culture

Indifference to history leading to a cult of modernity	Engineering style that seeks a technical fix
Impatience	Indifference to cultural distinctions
Continental outlook combined with a maritime situation and an airpower preference	Indifference to strategy
Resort to force characterized by a belated but massive response	Evasion of politics

Source: Adapted from Colin S. Gray, Explorations in Strategy, 89 – 98

Gray avers the nation's fascination with technology, which manifests itself as an engineering style seeking a technical fix, resulted from the conquest of the American West. The American frontier lacked a robust social support structure, which in turn "bred a pragmatism that translated into an engineering, problem-solving approach to life."[14] The sparse population density on the frontier and shortages of skilled laborers also drove a national preference for machines.[15] By the late nineteenth century, Americans tended to perceive new technologies both "as instruments of power and as triumphant symbols of human progress."[16] This belief in the power of technology has endured in American culture.

Particularly since the end of World War II, the American belief in the power of technology has affected how the nation wages war. As exemplified by the Manhattan Project, World War II required the "wholesale mobilization of American science and technology."[17] During the Cold War, the United States and the North Atlantic Treaty Organization (NATO) sought to employ a qualitative advantage arising

[14] Gray, *Explorations in Strategy*, 88.
[15] Gray, *Explorations in Strategy*, 88 – 89.
[16] Merritt Roe Smith, "Technological Determinism in American Culture," in *Does Technology Drive History? The Dilemma of Technological Determinism*, ed. Merritt Roe Smith and Leo Marx (Cambridge, MA: MIT, 1994), 8.
[17] Mahnken, *Technology and the American Way of War Since 1945*, 5.

from advanced technology to counterbalance the quantitative advantage of the Soviet Union and Warsaw Pact. Post-Cold War conflicts in Iraq, the former-Yugoslavia, and Afghanistan have continued to highlight the American use of advanced technologies, such as stealth and precision-guided munitions, in war.[18] Seeking a technological advantage in warfare, however, has a dark side. It reinforces a "dangerous American tendency" among policy-makers to "seek refuge in technology from hard problems of strategy and policy."[19]

The American penchant for employing technological solutions in war is neither positive nor negative. It does provide, however, the context for strategic action.[20] As Colin Gray argues,

> The machine-mindedness that is so prominent in the dominant American 'way of war' is inherently neither functional nor dysfunctional. When it inclines Americans to seek what amounts to a technological, rather than a political, peace, and when it is permitted to dictate tactics regardless of the political context, then on balance it is dysfunctional. Having said that, however, prudent and innovative exploitation of the technological dimension to strategy and war can be a vital asset.[21]

Assessing the impact of the 1991 Persian Gulf War, Eliot Cohen, Professor of Strategic Studies at Johns Hopkins University, argues air warfare is "distinctively American–high-tech, cheap in lives and (at least in theory) quick."[22] He cautioned, however, that the conflict's most dangerous legacy was the "fantasy of near bloodless uses of force."[23] These characteristics, Cohen argues, make airpower "an unusually seductive form of military strength" for American policy-makers.[24]

---

[18] Mahnken, *Technology and the American Way of War Since 1945*, 5.
[19] Weigley, *The American Way of War*, 416.
[20] Mahnken, *Technology and the American Way of War Since 1945*, 6.
[21] Gray, *Modern Strategy*, 147.
[22] Eliot Cohen, "The Mystique of US Air Power," *Foreign Affairs* 73, no. 1 (Jan – Feb 1994), 120.
[23] Cohen, "The Mystique of US Air Power," 121.
[24] Cohen, "The Mystique of US Air Power," 109.

Employed in this manner, airpower becomes a method through which policy-makers can conduct a strategy of coercive diplomacy. Coercion seeks to force an adversary to alter its behavior through the manipulation of costs and benefits.[25] Since a state may avoid the violence associated with coercion through accommodation, coercion is "the power to hurt," providing bargaining power for foreign policy.[26] As Thomas Schelling avers, "It is the *threat* of damage, or of more damage to come, that can make someone yield or comply. It is *latent* violence that can influence someone's choice–violence that can still be withheld or inflicted, or that a victim believes can be withheld or inflicted."[27] Thus, coercive diplomacy is the backing of a demand on an adversary "with a threat of punishment for noncompliance that [the enemy] will consider credible and potent enough to persuade him to comply with the demand."[28]

Coercive diplomacy, however, is a potentially "beguiling strategy."[29] It presents an attractive option for policy-makers seeking to achieve reasonable objectives in a crisis with lower political costs. Furthermore, coercive diplomacy appears to present a lower risk of unwanted escalation than traditional military action. The perceived efficacy of coercive diplomacy, though, can tempt policy-makers of powerful states into believing "that they can, with little risk, intimidate weaker opponents into giving up their challenge to a status quo situation."[30] If the interests involved strongly motivate the weaker state, however, then that state

---

[25] Robert Pape, *Bombing to Win: Air Power and Coercion in War* (Ithaca, NY: Cornell University, 1996), 4. The goal of both coercion and deterrence is to influence an adversary's behavior. While coercion seeks to force an adversary to alter behavior, deterrence aims to maintain the status quo by discouraging an adversary from taking certain actions.
[26] Thomas C. Schelling, *Arms and Influence*, 2008 ed. (New Haven, CT: Yale University, 2008), 2.
[27] Schelling, *Arms and Influence*, 3.
[28] Alexander L. George, *Forceful Persuasion: Coercive Diplomacy as an Alternative to War* (Washington, DC: United States Institute of Peace, 1994), 4.
[29] Alexander L. George and William E. Simons, eds, *The Limits of Coercive Diplomacy,* 2d Edition (Boulder, CO: Westview, 1994), 9.
[30] George and Simons, *The Limits of Coercive Diplomacy*, 9.

may call the bluff of the coercing power by refusing to back down. The stronger state must then decide whether to back down or escalate the confrontation.[31]

Employing force to achieve limited political objectives is problematic. Nuclear weapons only intensify this problem. President Eisenhower believed a war with the Soviet Union would automatically escalate into general nuclear war. Therefore, he developed a strategy to avoid nuclear war by making American military policy "so dangerous that his advisers would find it impossible to push Eisenhower toward war and away from compromise."[32] Eisenhower's approach to military policy eliminated limited war options by diminishing conventional capabilities at the expense of nuclear capabilities.[33] Upon taking office, however, President Kennedy and the members of his administration felt constrained by the defense policy and military force structure of the Eisenhower administration. The new president and his administration sought a broader "range of *usable* military power" to enable the nation to take the initiative in foreign policy.[34] Creating this capability required a stronger conventional military to allow fighting in limited, local wars.[35] This stronger conventional force, however, led to the nation's bitter experiences in Vietnam and Beirut.

Frustrated by these experiences, the Reagan administration re-evaluated the appropriate use of American military force, with Secretary

---

[31] George and Simons, *The Limits of Coercive Diplomacy*, 9. George argues seven factors favor, though do not guarantee, effective coercive diplomacy: (1) clarity of the objective, (2) strength of motivation, (3) asymmetry of motivation, (4) sense of urgency, (5) adequate domestic and international support, (6) opponent's fear of unacceptable escalation, and (7) clarity concerning the precise terms of settlement of the crisis. See George, *Forceful Persuasion*, 75 – 81. Pape is saturnine about the efficacy of coercion, arguing a state should attempt a strategy of coercion "only over issues so important to the coercer that it would be willing to pay the full costs of military victory….Coercion is no easier, only sometimes cheaper, and never much cheaper, than imposing demands by military victory." See Pape, *Bombing to Win*, 330 – 331.
[32] Campbell Craig, *Destroying the Village: Eisenhower and Thermonuclear War* (New York: Columbia University, 1998), 69.
[33] Craig, *Destroying the Village*, 84.
[34] Weigley, *The American Way of War*, 445.
[35] Weigley, *The American Way of War*, 445.

of Defense Casper Weinberger and Secretary of State George Schultz as the central protagonists. Secretary Weinberger's view that the nation should resort to force only when vital interests are at stake became the administration's view.[36] He announced the new American policy, eponymously known as the Weinberger Doctrine, at a 28 November 1984 speech before the National Press Club. In the speech, he provided six tests when weighing the use of military force.

> First, the United States should not commit forces to combat overseas unless the particular engagement or occasion is deemed vial to our national interest or that of our allies.
>
> Second, if we decide it is necessary to put combat troops into a given situation, we should do so wholeheartedly and with the clear intention of winning.
>
> Third, if we do decide to commit forces to combat overseas, we should have clearly defined political and military objectives…
>
> Fourth, the relationship between our objectives and the forces we have committed–their size, composition and disposition–must be continually reassessed and adjusted if necessary.
>
> Fifth, before the U.S. commits combat forces abroad, there must be some reasonable assurance we will have the support of the American people and their elected representatives in Congress…
>
> Finally, the commitment of U.S. forces to combat should be a last resort.[37]

This doctrine guided the use of force through the remainder of the Cold War and Desert Storm.

In the post-Cold War era, however, some American policy-makers believed this approach significantly restricted available policy options. Similar to the Kennedy administration, the Clinton administration felt

---

[36] Colin L. Powell, *My American Journey* (New York: Random House, 1995), 302 – 303.
[37] Casper Weinberger, "Excerpts from Address of Weinberger," *New York Times*, 29 November 1984.

constrained by the policies of previous administrations concerning the employment of conventional military force. Recognizing the continuing necessity for the use of force in the international system in the post-Cold War era, officials in the Clinton administration sought a way to make the use of force a viable option when non-vital interests were at stake.[38] This belief prompted then United States Ambassador to the United Nations Madeleine Albright to demand of General Colin Powell, "What's the point of having this superb military that you've always been talking about if we can't use it?"[39]

Advanced military technology provided a mechanism with which to solve this difficult problem. Such technology, particularly precision-guided munitions delivered from standoff platforms, presented a seductively alluring option for American policy-makers because it provided them with "a logical and innovative response" to both domestic and international constraints they face on the use of force.[40] Internationally, the focus on protecting civilians during armed conflict transformed efforts to limit collateral damage from "a vexing moral dilemma to a serious strategic problem."[41] Domestically, American policy-makers perceived the risking of American lives for policy aims short of serious threats to American national security as a politically difficult problem. Therefore, advanced technologies offering precision while concomitantly reducing risk to military personnel provided policy-makers with a mechanism to reconcile these competing interests for achieving policy aims when the use of force was necessary but the provocation fell short of a serious threat to national security.[42]

---

[38] Reuben E. Brigety II, *Ethics, Technology, and the American Way of War: Cruise Missiles and US Security Policy* (New York: Routledge, 2007), 4 – 5.
[39] Eric Schmitt, "The World; The Powell Doctrine is Looking Pretty Good Again," *The New York Times*, 4 April 1999, http://www.nytimes.com/1999/04/04/weekinreview/the-world-the-powell-doctrine-is-looking-pretty-good-again.html?pagewanted=all&src=pm (accessed 27 April 2012).
[40] Brigety, *Ethics, Technology, and the American Way of War*,1.
[41] Brigety, *Ethics, Technology, and the American Way of War*, 1.
[42] Brigety, *Ethics, Technology, and the American Way of War*,1 – 2.

In the 1990s, airpower frequently appeared to solve this strategic problem for American policy-makers. In fact, when coupled with precision guided munitions, airpower seemed to provide policy-makers "the ability to coerce Iraq, intervene in the Balkans, and retaliate against terrorist groups while avoiding the difficult decisions associated with a sustained commitment of ground forces."[43] The development of a system capable of an autonomous lethal engagement would provide policy-makers with yet another tool to circumvent the problems stemming from the competing interests of protecting civilians while also protecting American military personnel.

Defining and Measuring Autonomy

A system capable of conducting an autonomous lethal engagement would mark the next trend in battlefield automation. Automation on the battlefield is not a new development. In order to improve the accuracy of naval gunfire during World War I, officers sought to automate the complex calculations necessary to predict an enemy ship's location.[44] Between the World Wars, engineers continued work on battlefield automation through the development of aircraft bombsights and anti-aircraft artillery fire control systems.[45] The Norden bombsight, used extensively in World War II, was a critical step in the automation of an aircraft's bomb release. By the 1991 Persian Gulf War, automation of the bombing process reached the point where the B-52 navigation computer automatically opened the bomb bay doors and released the weapons at the correct point.[46]

Similar trends have occurred in other weapons systems. The Aegis combat system on US Navy *Ticonderoga*-class guided missile cruisers

[43] Mahnken, *Technology and the American Way of War Since 1945*, 179.
[44] David A. Mindell, *Between Human and Machine: Feedback, Control, and Computing before Cybernetics* (Baltimore, MD: Johns Hopkins University, 2002), 20 – 21.
[45] Mindell, *Between Human and Machine*, 43, 82
[46] Peter W. Singer, *Wired for War: The Robotics Revolution and Conflict in the 21st Century* (New York: Penguin, 2009), 124.

defends ships against airborne threats. Operators can set the Aegis system to operate in one of four modes with varying levels of autonomy ranging from direct human control to autonomous operation.[47] The Aegis system forms the basis of the US Army's Counter-Rocket, Artillery, and Mortar system, which protects bases against incoming threats.[48]

Describing a system as autonomous or automatic can create confusion about its capabilities. This confusion emerges because consensus does not exist on the definitions of automation and autonomy. The temptation arises to follow the example provided by United States Supreme Court Justice Potter Stewart with obscenity. When asked to provide a definition, he demurred and argued, "I know it when I see it."[49] One may say the same about automation and autonomy. In fact, National Aeronautics and Space Administration (NASA) researchers asked autonomous systems experts to provide an informal definition of autonomy. While the range of responses provided some "strong commonalities," several important aspects of the answers could not have been "more disjoint[ed]."[50] Various definitions of autonomy follow below:

- The "process by which essential functions can be performed with partial, intermittent, or no intervention from the operator."[51]
- The "execution by a machine agent (usually a computer) of a function that was previously carried out by a human."[52]

---

[47] Singer, *Wired for War*, 124. The four modes of operation are semiautomatic, in which the human operator determines which targets to engage; automatic special, in which the human operator designates priorities that the system then automatically engages; automatic, in which the system provides data to human operators but functions without them; and casualty, in which the system functions autonomously
[48] Singer, *Wired for War*, 124.
[49] *Jacobellis v. Ohio*, 378 U.S. 184 (1964).
[50] Johann Schumann and Willem Visser, "Autonomy Software: V&V Challenges and Characteristics," *2006 IEEE Aerospace Conference* (Big Sky, MT: IEEE, 4 – 11 March 2006), 2.
[51] USAF Test Pilot School, "Technology and Automation," *Systems Phase Text Book Chapter 3 – Human Factors* (Edwards AFB, CA: Air Force Material Command, July 2002), 3-37.
[52] Raja Parasuraman, "Humans and Automation: Use, Misuse, Disuse, Abuse," *Human Factors* 39, no.2 (June 1997), 231.

- The "capacity to operate in the real-world environment without any form of external control, once the machine is activated and at least in some areas of operations, for extended periods of time."[53]
- The "ability to generate one's own purposes without any instruction from outside…having free will."[54]
- A system's "own ability of integrated sensing, perceiving, analyzing, communicating, planning, decision-making, and acting/executing, to achieve its goals as assigned."[55]

These definitions provide a degree of commonality. Autonomy involves some degree of freedom or independence from a human operator and implies a level of adaptability to accomplish real-world tasks or goals.[56] Autonomy, however, is not an absolute, Manichean quality. Since it involves a degree of independence, autonomy exists along a spectrum of capabilities.

Autonomous and automatic capabilities, while typically used interchangeably, have subtle differences. An automatic system performs a specific, programmed task with minimal outside influence. An autonomous system, on the other hand, has some degree of say, or a type of free will, in task accomplishment. For example, while the common household dishwasher is automatic, few would describe it as autonomous. The *USAF UAS Flight Plan* echoes this delineation between automation and autonomy, asserting automation "differs from full

---

[53] Patrick Lin, George Bekey, and Keith Abney, *Autonomous Military Robotics: Risk, Ethics, and Design*, Office of Naval Research Report N00014-07-1-1152 (San Luis Obispo, CA: California Polytechnic State University, 20 December 2008), 105.
[54] Bruce Clough, "Metrics, Schmetrics! How the Heck do you Determine a UAV's Autonomy Anyway?" *2002 Performance Metrics for Intelligent Systems Workshop* (Gaithersburg, MD, 13 – 15 August 2002), 1.
[55] Hui-Min Huang, Elena Messina, and James Albus, NIST Special Publication 1011-II-1.0, *Autonomy Levels for Unmanned Systems (ALFUS) Framework Volume II: Framework Models Version 1.0* (Gaithersburg, MD: National Institute of Standards and Technology, December 2007), 16.
[56] Major Aaron M. Hopper, "The Future of Autonomy in U.S. Air Force Unmanned Air Systems: Toward a Strategy for Growth," (master's thesis, Air Command and Staff College, 2011), 14 – 15.

autonomy in that the system will follow preprogrammed decision logic."[57] To highlight the difference further, consider the following example. A basic aircraft autopilot is an automatic system to maintain a specific course chosen by the pilot. For the system to be autonomous, the guidance or navigation system would need to determine a particular course to take and then maintain the chosen course.[58] It is possible for a system to have both automatic and autonomous capabilities. Thus, a cruise missile that flies the route developed by mission programmers is automatic. This automatic system, however, would display a degree of autonomy if it had the capability to navigate around detected threats while executing its mission. Yet since autonomous systems do not possess the free will of a sentient being, they are following pre-programmed decision logic to some extent.[59] Therefore, the delineation between automation and autonomy breaks down to some degree, precipitating the need for a more nuanced method to describe a system's capabilities.[60]

In the early 2000s, researchers at the Air Force Research Laboratory (AFRL) led by Bruce Clough found describing the autonomy of an unmanned system to be a nebulous task. To remedy this, the research team developed autonomous control levels (ACL) metrics by utilizing John Boyd's Observe-Orient-Decide-Act (OODA) loop as a construct for measuring the effectiveness of autonomous software. The areas for analysis via the ACLs are Perception/Situational Awareness, Analysis/Coordination, Decision Making, and Capability, which correspond to the Observe, Orient, Decide, and Act steps of the OODA loop.[61] Table 2 presents the AFRL ACL metrics.

---

[57] Department of the Air Force, *United States Air Force Unmanned Aircraft Systems Flight Plan 2009 – 2047*, 33.
[58] Clough, "Metrics, Schmetrics!" 1.
[59] Lin, *Autonomous Military Robotics*, 104 – 105.
[60] Hopper, "The Future of Autonomy in U.S. Air Force Unmanned Air Systems," 18 – 19.
[61] Clough, "Metrics, Schmetrics!" 5 – 8.

Table 2: AFRL Autonomous Control Level Metrics

Level	Level Descriptor	Observe Perception/Situational Awareness	Orient Analysis/Coordination	Decide Decision Making	Act Capability
10	**Fully Autonomous**	Cognizant of all within Battlespace	Coordinates as necessary	Capable of total independence	Requires little guidance to do job
9	**Battlespace Swarm Cognizance**	Battlespace inference – intent of self and others (allies and foes) Complex/intense environment – on-board tracking	Strategic group goals assigned Enemy strategy inferred	Distributed tactical group planning Individual determination of tactical goal Individual task planning/execution Choose tactical targets	Group accomplishment of strategic goal with no supervisory assistance
8	**Battlespace Cognizance**	Proximity inference – intent of self and others (allies and foes) Reduced dependence upon off-board data	Strategic group goals assigned Enemy tactics inferred Automatic target recognition	Coordinated tactical group planning Individual task planning/execution Choose targets of opportunity	Group accomplishment of strategic goal with minimal supervisory assistance (example: go SCUD hunting)
7	**Battlespace Knowledge**	Short track awareness – History and predictive battlespace data in limited range, timeframe, and numbers Limited inference supplemented by off-board data	Tactical group goals assigned Enemy trajectory estimated	Individual task planning/execution to meet goals	Group accomplishment of tactical goal with minim supervisory assistance
6	**Real Time Multi-Vehicle Cooperation**	Ranged awareness – on-board sensing for long range, supplement by off-board-data	Tactical group goals assigned Enemy location sensed/estimated	Coordinated trajectory planning and execution to meet goals – group optimization	Group accomplishment of tactical goal with minimal supervisory assistance Possible close air space separation (1 – 100 yards)
5	**Real Time Multi-Vehicle Coordination**	Sensed awareness – Local sensors to detect others Fused with off-board data	Tactical group plan assigned RT Health Diagnosis; Ability to compensate for most failures and flight conditions; Ability to predict onset of failures (e.g. Prognostic Health Mgmt) Group diagnosis and resource management	On-board trajectory replanning – optimizes for current and predictive conditions Collision avoidance	Group accomplishment of tactical plan as externally assigned Air collision avoidance Possible close air space separation (1 – 100 yards) for AAR, formation in non-threat conditions
4	**Fault/Event Adaptive Vehicle**	Deliberate awareness – allies communicate data	Tactical plan assigned Assigned Rules of Engagement RT Health Diagnosis; Ability to compensate for most failures and flight conditions – inner loop changes reflected in outer loop performance	On-board trajectory replanning – event driven Self resource management Deconfliction	Self accomplishment of tactical plan as externally assigned Medium vehicle airspace separation (100's of yds)
3	**Robust Response to Real Time Faults/Events**	Health/status history and models	Tactical plan assigned RT Health Diag (What is the extent of the problems?) Ability to compensate for most control failures and flight conditions (i.e. adaptive inner-loop control)	Evaluate status vs. require mission capabilities Abort/RTB if insufficient	Self accomplishment of tactical plan as externally assigned
2	**Changeable Mission**	Health/status sensors	RT Health diagnosis (Do I have problems?)	Execute preprogrammed or updated plans in response to mission and health conditions	Self accomplishment of tactical plan as externally assigned
1	**Execute Preplanned Mission**	Preloaded mission data Flight Control and Navigation Sensing	Pre/Post BUT Report Status	Preprogrammed mission and abort plans	Wide airspace separation requirements (miles)
0	**Remotely Piloted Vehicle**	Flight Controls (attitude, rates) sensing Nose camera	Telemetered data Remote pilot commands	N/A	Control by remote pilot

Source: Bruce Clough, "Metrics, Schmetrics!" 8.

Plotting the ACLs on a radar chart, as shown in Figure 2 for a notional comparison of two autonomous systems, provides a visual representation of system characteristics and facilitates comparisons among different systems.[62]

[62] Clough, "Metrics, Schmetrics!" 5 – 8.

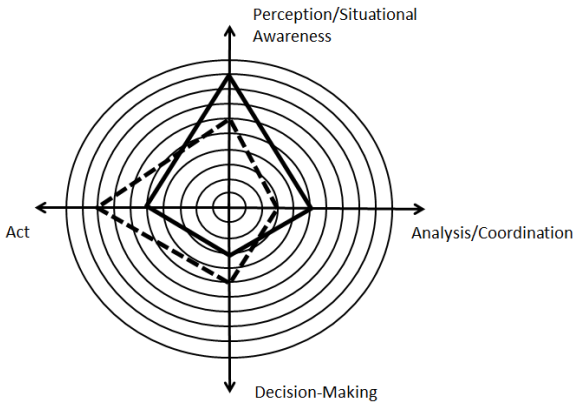

Figure 2: Comparison of Autonomous Systems Using AFRL ACL Metrics

Source: Adapted from Bruce Clough, "Metrics, Schmetrics!" 6.

Clough's team found these radar charts enabled management to quickly grasp the capability differences among various autonomous systems under development. These charts also aided the identification of capability gaps helping to focus the direction of technical research efforts.[63]

Building upon the work by AFRL and other stakeholders in the fields of robotics and autonomous systems, the National Institute of Standards and Technology (NIST) developed the autonomy levels for unmanned systems (ALFUS) framework. The ALFUS framework evaluates the contextual autonomous capability model of an unmanned system. NIST defines the contextual autonomous capability of an unmanned system as "characterized by the missions that the system is capable of performing, the environments within which the missions are performed, and human independence that can be allowed in the performance of the missions."[64] Through contextual autonomous capability, the ALFUS framework measures three specific aspects of an unmanned system: mission complexity, environmental complexity, and human

[63] Clough, "Metrics, Schmetrics!" 5 – 8.
[64] Huang, *ALFUS Framework*, 17.

independence. Plotting capabilities on these axes provides a method to compare system capabilities. Figure 3 illustrates the comparison of two notional unmanned systems on the NIST three-axis model for the ALFUS framework.

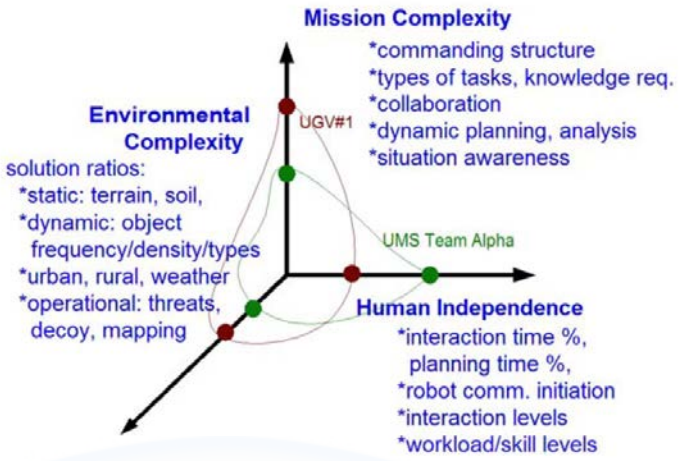

Figure 3: NIST Three-Axis Model for ALFUS

Source: Hui-Min Huang, ALFUS Framework, 23

Under the ALFUS framework, the autonomy level of an unmanned system evaluates the human independence axis, while the environmental complexity and mission complexity axes provide the context. As a general reference, the ALFUS framework defines three levels of contextual autonomous capability.

- Highest contextual autonomous capability–Completes all assigned missions with highest complexity; understands, adapts to, and maximizes benefit/value/efficiency while minimizing costs/risks on the broadest scope environmental and operational changes; capable of total independence from operator intervention.

- Mid contextual autonomous capability–Plans and executes tasks to complete an operator specified mission; limited understanding and response to environmental and operational changes and information; limited ability to reduce costs/risks while increasing benefit/value/efficiency; relies on about 50% operator input.

- Lowest contextual autonomous capability–Remote control for simple tasks in simple environments.[65]

Figure 4 provides a graphical representation of the spectrum of the ALFUS framework's contextual autonomous capability.

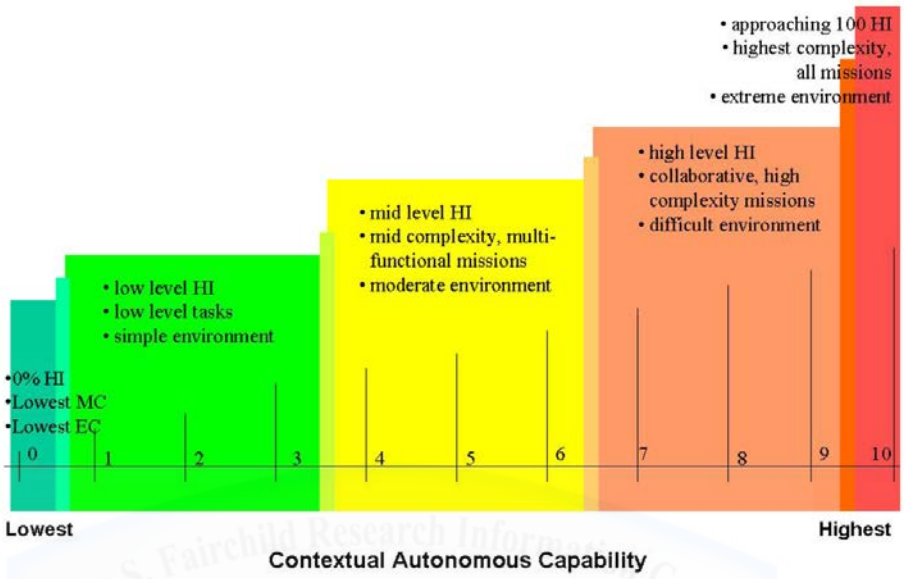

Figure 4: Illustration of ALFUS Contextual Autonomous Capability

Source: Hui-Min Huang, ALFUS Framework, 20

An overarching theme identified in the USAF's 2010 *Technology Horizons* to guide science and technology efforts is a transition "from control to autonomy."[66] As this transition occurs, the contextual autonomous capability of such systems will likely increase. The systems capable of autonomously employing lethal force considered in this work will reside at the higher levels of contextual autonomous capability.

Outline of the Investigation

Even though political and military leaders have not resolved legal, moral, and ethical questions surrounding the development of autonomous systems capable of employing lethal force, the development

[65] Huang, *ALFUS Framework*, 19 – 20.
[66] Werner J.A. Dahm, *Technology Horizons: A Vision for Air Force Science and Technology During 2010 – 2030*, AF/ST-TR-10-01-PR, (Washington, DC: Department of the Air Force, 15 May 2010), viii.

of these systems is nevertheless underway. Military professionals must understand these concerns in order to inform policy-makers as they attempt to resolve and reconcile these issues.

Chapter 2 will examine the legal and ethical concerns with autonomy. It will note that between the World Wars, the development of strategic bombardment doctrine occurred despite the lack of clear legal guidance with unresolved moral and ethical questions. An analogous situation regarding lethal autonomous systems exists today. The chapter will examine the law of armed conflict to discern principles affecting the development of lethal autonomous systems. It will then examine the moral and ethical arguments both for and against lethal autonomous systems.

Chapter 3 will explore the development of lethal autonomous systems. It will begin by examining the impact of science fiction in developing an ethical construct to guide the actions of robots. Next, it will investigate the top-down, bottom-up, and hybrid approaches scientists are analyzing as possibilities for implementing ethical reasoning in autonomous systems. The chapter will then look at the work of Ronald Arkin of the Georgia Institute of Technology and John Canning of the Naval Surface Warfare Division as specific approaches under development and assess the implications of these approaches. Finally, it will examine the verification and validation challenges that scientists must overcome to assuage concerns with lethal autonomous systems.

Chapter 4 will analyze human interaction with autonomous systems. It will begin by examining the emergence of human factors engineering as technological advancements blur the line between human and machine. The chapter will then delve into the resulting ethical and moral implications from human-machine interactions. Finally, it will address the importance of scrutinizing the moral and ethical consequences of developed advanced technology. The final chapter will provide a summary and concluding thoughts.

Chapter 1

Legal and Ethical Concerns with Autonomy

> *The passion for inflicting harm, the cruel thirst for vengeance, an unpacific and relentless spirit, the fever of revolt, the lust of power, and such like things, all these are rightly condemned in war.*
>
> — Saint Thomas Aquinas
> *The Summa Theologica, Part II, Question 40*

As Europe emerged from the ashes of World War I, the Italian airpower theorist Giulio Douhet believed the aircraft would change the fundamental character of warfare, leading future wars to be "total in character."[1] The airplane would enable military forces to attack an enemy's heartland without first breaking through the enemy's defenses. The enemy's citizens would become combatants since the aerial offensive would transform the entire enemy nation into a battlefield. Douhet's theories generated a paradox. While causing regrettable loss of life through attacks on an enemy's civilian population, aerial warfare could be a more humane form of war by causing hostilities to cease more quickly.

> Tragic, too, to think that the decision in this kind of war must depend upon smashing the material and moral resources of a people caught up in a frightful cataclysm which haunts them everywhere without cease [*sic*] until the final collapse of all social organization. Mercifully, the decision will be quick in this kind of war, since the decisive blows will be directed at civilians, that element of the countries at war least able to sustain them. These future wars may yet prove to be more humane than wars in the past in spite of all, because they may in the long run shed less blood.[2]

---

[1] Giulio Douhet, *The Command of the Air*, ed. Joseph Harahan and Richard Kohn (Tuscaloosa, AL: University of Alabama, 2009), 6.
[2] Douhet, *The Command of the Air*, 61.

Douhet thus accepted the loss of civilian life as part of a larger utilitarian calculation for the entire war.

American airmen confronted this moral grey area as they developed American airpower theory at the Air Corps Tactical School (ACTS) at Maxwell Field in Montgomery, AL. The 1935 ACTS text, *Air Force – Part I: Character and Strategy of Air Power*, asserted aerial attacks against an enemy's homeland would occur either to deny the enemy means for producing war material or to diminish civilian morale as part of an effort to have the population urge its government to cease hostilities.[3] ACTS argued direct attacks against civilians "cannot be justified on the basis of efficiency," yet conceded these attacks might nonetheless occur, necessitating planning for their eventual occurrence.[4]

During World War II, Allied commanders continued to struggle with the moral questions associated with attacks against civilian populations.[5] General Hap Arnold sent a memorandum to his combat commanders in June 1943 warning that Allied bombing of Germany would intensify hatred in the German populace against the Allies and thus poison the international environment after the war.[6] Calling for increased bombing accuracy, General Arnold reminded commanders that the Allied bombers, depending on employment tactics, could be either "the savior or the scourge of humanity."[7] General Arnold's concerns

---

[3] ACTS, "Character and Strategy of Air Power," *Air Force: Part One*, 1 Dec 1935, in AFHRC, decimal file no. 248.101-1, para. 28.
[4] ACTS, "Character and Strategy of Air Power," *Air Force: Part One*, 1 Dec 1935, in AFHRC, decimal file no. 248.101-1, para. 28.
[5] Ronald Schaffer argues practical concerns outweighed moral questions when Allied commander examined attacks against civilians. See Ronald Schaffer, "American Military Ethics in World War II: The Bombing of German Civilians," *The Journal of American History* 67, no. 2 (September 1980): 318 – 334. On page 319, he argues, "Official policy against indiscriminate bombing was so broadly interpreted and so frequently breached as to become almost meaningless…In the end, both the policy and apparent ethical support for it among AAF leaders turn out to be myths; while they contain elements of truth they are substantially fictitious or misleading."
[6] Ronald Schaffer, "American Military Ethics in World War II: The Bombing of German Civilians," *The Journal of American History* 67, no. 2 (September 1980), 333.
[7] General Hap Arnold, Memorandum to All Air Force Commanders in Combat Zones, 10 June 1943, quoted in Schaffer, "American Military Ethics in World War II," 333.

echoed those of Douhet and "represent a moral attitude inherent in air power theory… that bombing is a way of preserving lives by ending wars quickly and by providing a substitute for the kind of ground warfare that had killed so many soldiers a quarter century earlier."[8]

While struggling with the moral questions about the bombing of civilian populations, ACTS also examined the legal aspects. Despite the experiences during World War I, by the early 1930s nations had not agreed upon rules to govern aerial warfare. After examining the international legal landscape, ACTS concluded the rules of aerial warfare existed "only as a matter of form, not a single nation having ratified any of the several draft conventions proposed by various agencies, including the Hague, the International Law Association, the International Legal Committee on Aviation (*Comite Juridique de L'Aviation*) and the Red Cross."[9] The 1933-34 ACTS *International Aerial Regulations* text limned the state of rules governing aerial warfare by quoting Morton W. Royce, author of *Aerial Bombardment and the International Regulations of Warfare*, who asserted, "There are, thus, no conventional rules in actual force which directly effect [*sic*] aerial bombardment."[10]

As the potential emerges to enable machines to make lethal combat decisions, contemporary airmen find themselves in an ethical and legal situation similar to the one existing between the World Wars during the development of aerial warfare. Moral and ethical questions remain unresolved. Legal guidance is lacking. This chapter will examine the applicability of the existing international legal framework to lethal autonomous systems and capture the current ethics debates surrounding these systems.

---

[8] Schaffer, "American Military Ethics in World War II," 333.
[9] ACTS, "International Aerial Regulations" text, 1933-34, in AFHRC, decimal file no. 248.101-16, 24.
[10] ACTS, "International Aerial Regulations" text, 1933-34, in AFHRC, decimal file no. 248.101-16, 24.

Sources of International Law Governing Armed Conflict

International treaties and the United Nations (UN) Charter provide the foundation for the law governing armed conflict. The Hague Conventions of 1899 and 1907 and the Geneva Conventions of 1949 along with its Additional Protocols of 1977 and 2005 provide the legal foundations for the Law of Armed Conflict.[11] International treaties prohibiting the use of a specific type of weapon also contribute to the body of international law governing warfare. Two examples of these types of treaties are the 1997 Ottawa Treaty banning the use of anti-personnel land mines and the 2008 Cluster Munitions Convention.[12] International law regarding warfare aims to reduce the damage and suffering caused by war by restraining the actions of combatants.

The UN Charter provides international guidance concerning the role of force in the international systems and the conduct of belligerent nations. The purpose of the UN is to "maintain international peace and security" through "collective measures for the prevention and removal of threats to the peace."[13] Rather than resorting to force, the UN Charter urges member nations to "bring about by peaceful means, and in conformity with the principles of justice and international law, adjustment or settlement of international disputes or situations which might lead to a breach of the peace."[14] Chapter 1, Article 2, Section 4

---

[11] The Hague Conventions are available at http://www.icrc.org/ihl.nsf/385ec082b509e76c41256739003e636d/1d1726425f6955aec125641e0038bfd6 (accessed 23 January 2012). The Geneva Conventions are available at http://www.icrc.org/ihl.nsf/CONVPRES?OpenView (accessed 23 January 2012).
[12] The Ottawa Treaty is available at http://www.un.org/Depts/mine/UNDocs/ban_trty.htm (accessed 26 January 2012). The Cluster Munitions Convention is available at http://www.clusterconvention.org/files/2011/01/Convention-ENG.pdf (accessed 23 January 2012).
[13] United Nations, *Charter of the United Nations*, chapt. 1, art. 1, sec. 1, http://www.un.org/en/documents/charter/chapter1.shtml (accessed 10 January 2012).
[14] Charter of the United Nations, chapt. 1, art. 1, sec. 1.

discourages the threat or use of force in the international system.[15] Despite the proscriptions against the threat or use of force, Chapter 7, Article 51 allows for self-defense. In the event of an attack against a member nation, the Charter directs member nations to refer the matter to the UN Security Council, which should take the necessary steps to maintain international peace and security. Under Article 51, the attacked nation, however, has the authority to take appropriate defensive matters during Security Council deliberations.[16] Article 51, therefore, subsumes the use of force in the international system under certain circumstances and thereby makes the use of force lawful.[17]

While the UN Charter allows for the use of force in the international system under limited circumstances, it does not prescribe the rules governing the use of force. In the mid-1990s, the International Court of Justice, pursuant to UN General Assembly Resolution 49/75 K, examined the question, "Is the threat or use of nuclear weapons in any circumstance permitted under international law?"[18] Although the case examines the specific question of nuclear weapons, it provides insight into questions regarding the rules governing the use of force. The Court noted that the protections of the International Covenant on Civil and Political Rights, which protects a person's right to not arbitrarily be

---

[15] Charter of the United Nations, chapt. 1, art. 2, sec. 4. "All Members shall refrain in their international relations from the threat or use of force against the territorial integrity or political independence of any state, or in any other manner inconsistent with the Purposes of the United Nations."
[16] Charter of the United Nations, chapt. 7, art. 51. "Nothing in the present Charter shall impair the inherent right of individual or collective self-defense if an armed attack occurs against a Member of the United Nations, until the Security Council has taken measures necessary to maintain international peace and security. Measures taken by Members in the exercise of this right of self-defense shall be immediately reported to the Security Council and shall not in any way affect the authority and responsibility of the Security Council under the present Charter to take at any time such action as it deems necessary in order to maintain or restore international peace and security."
[17] Lt Col Chris Jenks, "Law from Above: Unmanned Aerial Systems, Use of Force, and the Law of Armed Conflict," *North Dakota Law Review* 85, no. 3 (2009), 658.
[18] *Legality of the Threat or Use of Nuclear Weapons*, Advisory Opinion, 1996 International Court of Justice 240 (8 July), 6.

deprived of life, did not cease during times of war.[19] The Court went on to argue, "The test of what is an arbitrary deprivation of life, however, then falls to be determined by the applicable *lex specialis,* namely, the law applicable in armed conflict which is designed to regulate the conduct of hostilities."[20] Furthermore, the Court averred that the law applicable in armed conflict should determine whether a loss of life caused by a certain type of weapon is an arbitrary deprivation of life.[21] This ruling, therefore, refers the question of legality for a specific weapon, such as an autonomous lethal system, back to the law of armed conflict.

The International Criminal Tribune for the former Yugoslavia also provides insight into the use of force in the international system. The *Prosecutor v. Tadic* decision informs the determination of when an international conflict rises to the level of armed conflict and of the applicability of the laws of armed conflict to such conflicts.[22]

> Armed conflict exists whenever there is a resort to armed force between States or protracted armed violence between governmental authorities and organized armed groups or between such groups within a State. International humanitarian law applies from the initiation of such armed conflicts and extends beyond the cessation of hostilities until a general conclusion of peace is reached; or, in the case of internal conflicts, a peaceful settlement is achieved. Until that moment, international humanitarian law continues to apply in the whole territory of the warring States or, in the case of internal conflicts, the whole territory under the

---

[19] *Legality of the Threat or Use of Nuclear Weapons*, Advisory Opinion, 1996 International Court of Justice 240 (8 July), 18 (paragraph 25).
[20] *Legality of the Threat or Use of Nuclear Weapons*, Advisory Opinion, 1996 International Court of Justice 240 (8 July), 18 (paragraph 25). *Lex specialis* is Latin for "law governing a specific subject matter." According to the legal doctrine *lex specialis derogate legi generali*, law governing a specific subject matter overrides law governing general matters. See "Lex Specialis Law and Legal Definition" at http://definitions.uslegal.com/l/lex-specialis/ (accessed 15 February 2012).
[21] *Legality of the Threat or Use of Nuclear Weapons*, Advisory Opinion, 1996 International Court of Justice 240 (8 July), 18 (paragraph 25).
[22] Lt Col Chris Jenks, "Law from Above: Unmanned Aerial Systems, Use of Force, and the Law of Armed Conflict," *North Dakota Law Review* 85, no. 3 (2009), 663.

> control of a party, whether or not actual combat takes place there.[23]

The court defines international humanitarian law as the four Geneva Conventions.[24] Using the *Tadic* case as a guide, when an armed conflict exists, "the full panoply of customary international law of International Humanitarian Law (IHL) applies to the *jus in bello* analysis."[25] Questions regarding the legality of lethal autonomous systems must then turn to the law of armed conflict.

Law of Armed Conflict Principles

Just War Theory provides a philosophical foundation for moral conduct in war by addressing both *jus ad bellum*, the justice of war, and *jus in bello*, justice in war. *Jus ad bellum* is concerned with when it is morally permissible to conduct warfare, while *jus in bello* addresses moral actions within the context of war. The Law of Armed Conflict provides the legal framework codifying *jus in bello*.[26]

The consensus on which military practices and weapons international law should prohibit has changed over time, yet the underlying principles of *jus in bello* have remained the same. Technology, along with society's attitude toward warfare, helps explain this changing

---

[23] Prosecutor v. Tadic, Case No. IT-94-1-I, Decision on the Defense Motion for Interlocatory Appeal on Jurisdiction, (2 October 1995), http://www.icty.org/x/cases/tadic/acdec/en/51002.htm (accessed 10 January 2012).
[24] Prosecutor v. Tadic, Case No. IT-94-1-I, Decision on the Defense Motion for Interlocatory Appeal on Jurisdiction, (2 October 1995), para. 67. The four Geneva Conventions are the Convention for the Amelioration of Condition of the Wounded and Sick in Armed Forces in the Field, 12 August 1949, 75 U.N.T.S. 970; the Convention for the Amelioration of the Condition of the Wounded, Sick, and Shipwrecked Members of the Armed Forces at Sea, 12 August 1949, 75 U.N.T.S. 971; Convention Relative to the Treatment of Prisoners of War, 12 August 1949, 75 U.N.T.S. 972; and Convention Relative to the Protection of Civilian Persons in Time of War, 12 August 1949, 75 U.N.T.S. 973.
[25] Jenks, "Law from Above," 663.
[26] Michael Walzer, *Just and Unjust Wars: A Moral Argument with Historical Illustrations* (New York: Basic Books, 1977), 21.

consensus.[27] For example, the bombing of cities during World War II was widely accomplished but not prosecuted as a war crime after the war. Planners clearly knew the bombing of cities to destroy the enemy's means of production would concomitantly create civilian casualties, yet Allied commanders partially defended this practice based on the technology available at the time.[28] However, as precision weapons technology has emerged, society has become less tolerant of civilian casualties.[29] Contemporary society would most likely not accept a bombing campaign against cities, such as the February 1945 bombing of Dresden.[30] Similarly, consensus on the use of lethal autonomous systems may evolve over time as the technology demonstrates, or fails to demonstrate, its efficacy.

The *lex specialis* international law governing warfare does not provide definitive guidance concerning the legality of autonomous weapons. The international community adopted The Hague and Geneva Conventions well before the potential of an autonomous weapon system existed. To determine the legality of autonomous weapon systems, therefore, a military must examine the underlying principles of the law of armed conflict.

---

[27] Armin Krishnan, *Killer Robots: Legality and Ethicality of Autonomous Weapons*, (Burlington, VT: Ashgate, 2009), 90.

[28] Bomber Command eventually determined crews could only effectively fly at night and that with available technology at night, only one in three bombers would drop its bombs within 5 miles of the desired aimpoint. At that time, the bombings could have been justified under the Just War principle of "Supreme Emergency." For further reading, see Walzer pages 255 to 268. For a further account of the development of World War II strategic bombing doctrine, see Tami Davis, *Rhetoric and Reality in Air Warfare: The Evolution of British and American Ideas About Strategic Bombing, 1914 – 1945* (Princeton, NJ: Princeton University, 2002).

[29] Krishnan, *Killer Robots*, 90.

[30] American commanders feared, even in the context of World War II, the American public might not accept terror bombing. When the Associated Press reported the decision to firebomb Dresden was a "long-awaited decision to adopt deliberate terror bombing…as a ruthless expedient to hasten Hitler's doom," American commanders were quick to correct the story and argue precision bombing remained American policy. For more information, see Michael Sherry, *The Rise of American Air Power: The Creation of Armageddon* (New Haven, CT: Yale University, 1987), 260 to 264 and Tami Biddle, *Rhetoric and Reality*, (Princeton, NJ: Princeton University), 254 to 257.

The four fundamental principles of necessity, proportionality, discrimination, and humaneness guide the application of force during war and form the basis for the law of armed conflict.[31] The principle of military necessity authorizes the use of military force against belligerents required to accomplish a mission.[32] The principle of necessity stems from the Hague Convention (IV) Respecting the Laws and Customs of War on Land and its Annex: Regulations Concerning the Laws and Customs of War on Land. The preamble states nations adopted this convention based on "the desire to diminish the evils of war, *as far as military requirements permit.*"[33] Article 23 of the Hague Convention (IV) provides a further foundation for the principle of necessity through its proscription of the destruction or seizure of the enemy's property unless "*imperatively demanded by the necessities of war.*"[34] The principle of necessity also requires a balancing test to ascertain the legality of a weapon. Before a weapon enters a military's inventory, a legal review must examine the weapons effects "against comparable weapons in contemporary use, their effects on combatants, *and* the military necessity for the weapon under consideration."[35]

According to the principle of proportionality, a nation's use of military force should be commensurate with military objectives in an effort to protect civilian populations and enemy combatants from

---

[31] *Air Force Operations and the Law: A Guide for Air, Space & Cyber Forces*, (Maxwell AFB, AL: The Judge Advocate General's School, 2009), 13. Discrimination is also referred to as distinction, while humaneness is also known as unnecessary suffering.
[32] *Air Force Operations and the Law*, 13.
[33] Preamble to the Hague Convention (IV) respecting the Laws and Customs of War on Land and its Annex: Regulations Concerning the Laws and Customs of War on Land. The Hague, 18 October 1907. (http://www.icrc.org/ihl.nsf/FULL/195?OpenDocument, accessed 23 January 2012), emphasis added.
[34] Article 23, Section (g), Hague Convention (IV) respecting the Laws and Customs of War on Land and its Annex: Regulations Concerning the Laws and Customs of War on Land. The Hague, 18 October 1907.
(http://www.icrc.org/ihl.nsf/FULL/195?OpenDocument, accessed 23 January 2012). The rhetoric of restrain and reality of total war do not always agree as exemplified by General William T. Sherman's adherence (or lack thereof) to the Lieber Code during the March to the Sea.
[35] *Air Force Operations and the Law*, 16.

excessive and unnecessary use of force. Proportionality weighs the anticipated gains of military operations against reasonably foreseeable consequences to the civilian population. Proportionality also provides the fulcrum for balancing unnecessary suffering and military necessity. The principle of proportionality requires the military commander to assess whether the incidental loss of civilian life, injury to civilians, and damage to civilian property would be excessive when compared to the military advantage anticipated from the operation. Decision-makers at all levels–strategic, operational, and tactical–must consider the principle of proportionality in conducting operations.[36]

The principle of discrimination requires belligerents to apply military force selectively and only against military targets, again in an effort to protect the civilian population from the use of force. The United States has codified the principle of discrimination (or distinction) in the conduct of its military since as early as the Civil War. The Lieber Code of April 1863 required the Union Army to spare unarmed citizens "in person, property, and honor as much as the exigencies of war will admit."[37] UN General Assembly Resolution 2444, passed on 19 December 1968, promoted the principle of distinction before completion of Additional Protocol I to the Geneva Convention (AP I). Resolution 2444 affirms distinction "must be made at all times between persons taking part in the hostilities and members of the civilian population to the effect that the latter be spared as much as possible."[38]

Article 48 of AP I establishes the basic principle of distinction. This article requires belligerents to "distinguish between the civilian population and combatants and between civilian objects and military objectives and accordingly shall direct their operations only against

---

[36] *Air Force Operations and the Law*, 19.
[37] Instructions for the Government of Armies of the United States in the Field (Lieber Code), 24 April 1863 (http://www.icrc.org/ihl.nsf/FULL/110?OpenDocument, accessed 23 January 2012).
[38] UN General Assembly Resolution 2444, 18 December 1968, (http://www.icrc.org/ihl.nsf/FULL/440?OpenDocument, accessed 23 January 2012).

military objectives."[39] Paragraphs 4 and 5 of AP I Article 51 limn the types of attacks considered indiscriminate, while paragraph 1 of AP I Article 57 directs belligerents to take "constant care…to spare the civilian population, civilians, and civilian objects."[40]

The principle of distinction contributes to the doctrine of double effect.[41] The doctrine emerged from the writings of Saint Thomas Aquinas in the *Summa Theologica.*[42] In considering whether it is lawful to kill a person in self-defense, Aquinas argues, "Nothing hinders one act from having two effects, only one of which is intended, while the other is beside the intention…Accordingly the act of self-defense may have two effects, one is the saving of one's life, the other is the slaying of the aggressor."[43] Double effect reconciles the "prohibition against attacking

---

[39] Article 48, Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts (Protocol I), 8 June 1977, 1125 U.N.T.S. 25, (http://treaties.un.org/doc/Publication/UNTS/Volume%201125/v1125.pdf, accessed 26 January 2012 or http://www.icrc.org/ihl.nsf/FULL/470?OpenDocument, accessed 23 January 2012). While the United States is a signatory to both Additional Protocol I and Additional Protocol II, the United States has not acceded to either treaty. The United States does, however, recognize aspects of both treaties as customary international law. For additional information, see note 76 in Jenks, "Law from Above," 665.

[40] Article 51, para. 5(b), Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I), 8 June 1977, 1125 U.N.T.S. 26. "5. Among others, the following types of attacks are to be considered as indiscriminate: (a) an attack by bombardment by any methods or means which treats as a single military objective a number of clearly separated and distinct military objectives located in a city, town, village or other area containing a similar concentration of civilians or civilian objects; and (b) an attack which may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage anticipated." Article 57, para. 1, Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I), 8 June 1977, 1125 U.N.T.S. 29.

[41] The doctrine of double effect is a moral argument one invokes to explain the permissibility of an action causing serious harm, such as death, as a side effect of promoting a good end. See Alison McIntyre, "Doctrine of Double Effect," *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition),* ed. Edward N. Zalta, http://plato.stanford.edu/archives/fall2008/entries/double-effect (accessed21 February 2012).

[42] McIntyre, "Doctrine of Double Effect.

[43] St. Thomas Acquinas, "Second Part of the Second Part, Question 64. Murder, Article 7. Whether it is lawful to kill a man in self-defense?" *The Summa Theologica*, 2008 on-

noncombatants with the legitimate conduct of military activity."[44] Princeton Professor Michael Walzer provides four conditions for the doctrine of double effect to hold.

1. The act is good in itself or at least indifferent, which means, for our purposes, that it is a legitimate act of war.
2. The direct effect is morally acceptable–the destruction of military supplies, for example, or the killing of enemy soldiers.
3. The intention of the actor is good, that is, he aims only at the acceptable effect; the evil effect is not one of his ends, nor is it a means to his ends.
4. The good effect is sufficiently good to compensate for allowing the evil effect; it must be justifiable under Sidgwich's proportionality rule.[45]

Walzer argues soldiers should "minimize the dangers they impose" and take "due care" to avoid harming civilians.[46] For Walzer, soldiers must assume mission risk to avoid killing civilians. The limit of these risks arises at the point where additional risk-taking would doom the mission or make it so costly that the military could not re-attempt the mission.[47] From this argument emerges the principle of double intention.

Finally, under the principle of humaneness, belligerents should apply no more military force than necessary to win the conflict and should avoid inflicting unnecessary suffering. Article 23 of the Hague Convention establishes the principle of humanness by forbidding certain acts, such as the employment of poisoned weapons, the declaration of no quarter, and the employment of "arms, projectiles, or material calculated to cause unnecessary suffering."[48] Article 23 of the Hague Convention also encodes aspects of chivalry by forbidding acts of perfidy–a hostile act

---

line edition, ed. Kevin Knight, http://www.newadvent.org/summa/3064.htm (accessed 21 February 2012).

[44] Walzer, *Just and Unjust Wars*, 153.

[45] Walzer, *Just and Unjust Wars*, 153.

[46] Walzer, *Just and Unjust Wars*, 157.

[47] Walzer, *Just and Unjust Wars*, 157. To illustrate this point, Walzer cites the example of Free French Air Force pilots bombing France who accepted greater personal risk to ensure precision bombing by flying at very low altitudes.

[48] Article 23, Hague Convention (IV) respecting the Laws and Customs of War on Land and its Annex: Regulations Concerning the Laws and Customs of War on Land. The Hague, 18 October 1907.

taken under the cover of legal protection.[49] Acts of perfidy forbidden in Article 23 include the killing of an enemy who has surrendered and the improper use of a flag of truce or of military insignia.[50]

The principle of humaneness has provided the foundation for international treaties, such as the Ottawa Treaty, that outlaw particular types of weapons or military practices.[51] While a specific treaty banning autonomous systems is not in force, leading critics against robotic military systems have formed the International Committee for Robot Arms Control (ICRAC).[52] ICRAC sees an "inexorable drive" to create lethal autonomous robots.[53] Troubled by the potential consequences of these systems, ICRAC is calling upon the international community to develop an arms control regime against these systems. Specifically, ICRAC calls for a prohibition on the development, deployment, and use of armed unmanned systems along with a ban on the arming of unmanned systems with nuclear weapons.[54]

Since a treaty proscribing the use of lethal autonomous systems does not exist, the military must assess these weapons against the law of armed conflict principles to ascertain their legality.[55] Lethal autonomous weapons do not appear to be inherently illegal based on the law of armed conflict principles, but the Devil remains in the details of implementation. An autonomous weapon system designed to maximize suffering or to apply force indiscriminately would clearly be illegal based

---

[49] For additional information on chivalry and perfidy, see *Air Force Operations and the Law*, 21.
[50] Article 23, Hague Convention (IV) respecting the Laws and Customs of War on Land and its Annex: Regulations Concerning the Laws and Customs of War on Land. The Hague, 18 October 1907.
[51] Krishnan, *Killer Robots*, 90 – 95.
[52] The founding members of ICRAC are Noel Sharkey, University of Sheffield; Robert Sparrow, Monash University; Jürgen Altmann, Dortmund Technische Universität; and Peter M. Asaro. See International Committee for Robot Arms Control (ICRAC), www.icrac.co.uk (accessed 17 February 2012).
[53] ICRAC, www.icrac.co.uk (accessed 17 February 2012).
[54] ICRAC, "Mission Statement," www.icrac.co.uk (accessed 17 February 2012).
[55] Article 36, Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts (Protocol I), 8 June 1977, 1125 U.N.T.S. 25.

on the law of armed conflict principles.[56] It may be possible, however, for autonomous weapon systems to be designed within the existing law of armed conflict framework.

Philosophers and ethicists disagree whether a lethal autonomous weapon system would violate the fundamental principles of the law of armed conflict. Under necessity, the use of an autonomous lethal weapon would become a military necessity when the autonomous weapon system is more effective than other available weapons, possibly to the extent that human soldiers become militarily ineffective.[57] If only autonomous machines fought battles, the lives of human soldiers would not be at risk. Some may see this as a positive development. Others would argue, however, that since human lives would not be at risk in battles fought entirely by autonomous machines, then the restraints on the use of force may be lowered. Consequently, more conflict would occur.[58]

Autonomous weapon systems have the technological potential to exercise more proportionality than human soldiers do.[59] Emotions would not cloud the decisions made by autonomous weapons, as can be the case with human soldiers who can experience fear and hatred during a battle. It is possible to envision an autonomous weapon system capable of calculating the advantages of the application of military force in a given circumstance and comparing it to the humanitarian cost more effectively and efficiently than a human soldier. Autonomous weapon systems, therefore, could theoretically achieve more proportionality in the application of lethal force.[60]

The principle of discrimination could be a limiting factor for fielding a lethal autonomous weapon system. The sophisticated sensors and precision weapons available to autonomous weapon systems could

---

[56] Krishnan, *Killer Robots*, 97.
[57] Krishnan, *Killer Robots*, 91 – 92.
[58] Krishnan, *Killer Robots*, 92.
[59] Krishnan, *Killer Robots*, 92.
[60] Krishnan, *Killer Robots*, 92.

enable these systems to apply force against enemy forces with more proportionality and more precisely than human soldiers. The ability to distinguish between a civilian and an enemy combatant, however, may prove to be a significant technical challenge for these systems. While the deliberate misuse of protected civilian sites, such as churches or hospitals, by enemy forces could pose problems for autonomous weapon systems, these systems could be designed to recognize such violations, report them, and engage forces under certain, authorized circumstances. Systems able to apply force under the principle of proportionality should also meet the humaneness requirement.[61] Having examined the legal foundation for assessing lethal autonomous weapon systems, attention now turns toward examining arguments against and for the use of these systems.

Arguments Against Autonomy: Discrimination and Proportionality

Noel Sharkey, a Professor of Artificial Intelligence at the University of Sheffield, is a critic of autonomous weapon systems. Sharkey argues autonomous weapon systems would violate the law of armed conflict principles of discrimination and proportionality.[62] He asserts, "No autonomous robots or artificial intelligence systems have the necessary skills to discriminate between combatants and innocents… There are no visual or sensing systems up to that challenge."[63] Sharkey also argues that even defining "civilian" for the system would be difficult since the Geneva Conventions effectively define a civilian as someone who is not a combatant.[64] Even if this problem were overcome, Sharkey avers the sensors on the autonomous system would only provide enough data to

[61] Krishnan, *Killer Robots*, 93 – 97.
[62] Noel Sharkey, "Cassandra or False Prophet of Doom: AI Robots and War," *IEEE Intelligent Systems,* July/August 2008, 16 – 17.
[63] Noel Sharkey, "Grounds for Discrimination: Autonomous Robot Weapons," *RUSI Defense Systems – Challenges of Autonomous Weapons*, October 2008, 87, http://www.rusi.org/downloads/assets/23sharkey.pdf (accessed 17 February 2012).
[64] Sharkey, "Grounds for Discrimination," 87 – 88.

indicate the presence of a human without making the distinction between combatant and civilian. Sharkey argues the system's computational power would not be sufficient to determine proportionality. He further argues programmers would not be able to quantify suffering, because a "known metric to objectively measure needless, superfluous or disproportionate suffering" does not exist.[65]

To illustrate his arguments, Sharkey points to the BLU-108 sub-munition used in CBU-97 Sensor Fused Weapons. The BLU-108 sub-munitions have infrared sensors to detect armored targets.[66] Sharkey argues that the sub-munitions do not have the ability to discriminate between armored targets and civilian vehicles, such as buses or cars.[67] The human employing the weapon must therefore discriminate between military target and civilian vehicles, thus employing the weapon appropriately.[68] Sharkey carries this analogy to autonomous weapon systems, arguing these weapons would not have a means to discriminate between military and civilian targets.[69] Sharkey's argument on the current inability of sensors to discriminate between military and civilian targets, however, falls into a line of reasoning against technological development whereby critics argue that current technological limitations are inherent and insurmountable.[70]

---

[65] Sharkey, "Grounds for Discrimination," 88.
[66] TEXTRON Systems, "BLU-108 Submuntion," http://www.textrondefense.com/assets/pdfs/datasheets/blu108_datasheet.pdf (accessed 15 February 2012).
[67] Sharkey, "Grounds for Discrimination," 89. According to the TEXTRON Systems BLU-108 datasheet, the warhead utilizes a two-color infrared sensor to detect targets matching specific infrared requirements. After the system validates a target, the explosives detonate. The TEXTRON datasheet does not describe the target validation process, which is most likely proprietary and classified. While the sophistication of the target validation process has the potential to obviate Sharkey's argument, his assertion on the system's inability to distinguish between an armored target and a lorry will be accepted to connect to his larger point about the human having to discriminate targets with many weapons.
[68] Sharkey, "Grounds for Discrimination," 89.
[69] Sharkey, "Grounds for Discrimination," 89.
[70] Ray Kurzweil, *The Singularity is Near: When Humans Transcend Biology* (New York: Viking, 2005), 240.

Arguments Against Autonomy: Accountability

Robert Sparrow, a Professor at Monash University in Victoria, Australia, argues employing lethal autonomous machines will be unethical unless someone could be held responsible for the lethal actions taken by the machine.[71] Sparrow examines the possibility of holding three entities responsible for the machine's actions: the manufacturer and/or designer, the commander authorizing the use of the lethal autonomous system, and the autonomous system itself.[72] Sparrow concludes accountability could only fall on the manufacturer and/or designer in instances of negligence.[73] He argues it would be unfair to hold commanders responsible for the actions of an autonomous system not completely under their control.[74] While Sparrow concludes current autonomous systems are not advanced enough to be held accountable for their actions, he envisions a future with autonomous systems having advanced artificial intelligence sufficient to make the machine a moral agent capable of being held accountable for its actions.[75] Most people, however, would balk at the idea of holding a machine morally responsible for its actions due to the difficulty of imagining how to hold a machine accountable.[76] Accountability implies the capacity for understanding punishment. Sparrow argues an artificial intelligence advanced enough to actually suffer and be held accountable for its actions would be equivalent to a full moral person. This status would then turn accountability back to humans, because humans could then be held accountable for their actions concerning the machine.[77] In this instance, humans would then be as "morally concerned when our machines are

[71] Robert Sparrow, "Killer Robots," *Journal of Applied Philosophy* 24, no. 1 (February 2007), 67.
[72] Sparrow, "Killer Robots," 68 – 71.
[73] Sparrow, "Killer Robots," 69 – 70.
[74] Sparrow, "Killer Robots," 70 – 71.
[75] Sparrow, "Killer Robots," 71.
[76] Sparrow, "Killer Robots," 72.
[77] Sparrow, "Killer Robots," 73.

destroyed–indeed killed–as we are when human soldiers die in war."[78] Due to the resulting paradox, the development of these advanced machines would not achieve the goal Sparrow sees as motivating these machines in the first place–enabling wars to be fought without risking the death of soldiers. Sparrow illustrates the accountability dilemma of using lethal autonomous systems by comparing these systems to using child soldiers.[79] Child soldiers are autonomous, yet they lack full moral autonomy, and therefore are not morally responsible for their actions. The inability to fairly hold a party responsible for the actions of an autonomous system makes the use of these systems unethical to Sparrow.[80]

While critics such as Sparrow argue that autonomous systems would lack battlefield accountability, it can still exist. In an instance of the improper utilization of an autonomous system, accountability for those actions would fall back on the military commander, as the military holds commanders accountable for the actions of their subordinates.[81] On the other hand, if an autonomous system does not operate as designed, then accountability would fall to the manufacturer.[82] A case where an autonomous system causes an incident not attributable to an improper design or improper use may stress the legal system. In this instance, the particular systems, and those of the same design, would need to be withdrawn.[83] If the system were not withdrawn in this particular case, "it can only be interpreted as a failure of politics and maybe as a war crime or crime against humanity committed by the political leadership of a state."[84]

---

[78] Sparrow, "Killer Robots," 73.
[79] Sparrow, "Killer Robots," 73.
[80] Sparrow, "Killer Robots," 62 – 75.
[81] Krishnan, *Killer Robots*, 104.
[82] Krishnan, *Killer Robots*, 105.
[83] Krishnan, *Killer Robots*, 105.
[84] Krishnan, *Killer Robots*, 105.

Regarding Sparrow's comparison between child soldiers and autonomous systems, Ronald Arkin, a robotics professor at the Georgia Institute of Technology, argues Sparrow "neglects, however, to consider the possibility of the direct encoding of prescriptive ethical codes within the robot itself, which can govern its actions in a manner consistent with the Laws of War and rules of engagement."[85] According to Arkin, the inclusion of this ethical coding would therefore weaken Sparrow's claims about the lack of moral accountability for the autonomous system.

Arguments Against Autonomy: Ease of Warfare

Peter Asaro from Rutgers University argues, "One of the strongest moral aversions to the development of robotic [systems] stems from the fear that they will make it easier for leaders to take an unwilling nation into war."[86] To help illustrate this point, Asaro cites the 1991 Persian Gulf War, the 1999 war in Kosovo, and the 2003 invasion of Iraq as examples where political leaders led the populace to believe military action would result in few friendly casualties, thus making the military action "safe." The development and use of autonomous systems, therefore, would limit the battlefield risks to soldiers and thus lower the political threshold for military action.[87]

George Mason University Professor of Government and Politics Reuben Brigety II notes an analogous trend in the American use of Tomahawk cruise missiles during the 1990s. In the post-Cold War era, both the domestic and international political environment constrained the ability of the United States to respond to provocations to American

[85] Ronald C. Arkin, *Governing Lethal Behavior in Autonomous Robots* (New York: CRC Press, 2009), 38.
[86] Peter Asaro, "How Just Could a Robot War Be?" in *Current Issues in Computing and Philosophy*, ed. Adam Briggle, Katinka Waelbers, and Philip Brey, (Fairfax, VA: IOS Press, 2008), 56.
[87] Asaro, "How Just Could a Robot War Be?" 56 – 58.

interests with massive military force.[88] Yet despite the constraints, American policy-makers perceived a necessity to use force under certain conditions. Thus the challenge American policy-makers faced was finding a feasible option to apply force that, to the extent possible, served American interests, protected civilians on the battlefield, and limited risks to American and allied soldiers. The Tomahawk cruise missile provided a technological solution, making military force a "viable instrument of statecraft in this contested environment."[89] Rather than viewing domestic constraints on the use of force stemming from the potential for military casualties as the result of a fickle public or wobbly politicians troubled by the realities of war, Brigety views this constraint as a form of democratic accountability for the commitment of military forces to combat.[90]

Sharkey also argues against autonomy based on the potential that autonomous weapons systems would lower the threshold for military conflicts. The use of autonomous systems would lower the risk to human soldiers, thus lowering casualties for the side utilizing autonomous systems. A lack of casualties would provide political leaders with fewer disincentives to initiate conflicts.[91]

Michael Ignatieff, a Canadian historian and politician, also makes a similar argument. Ignatieff describes NATO's use of force in Kosovo as "virtual war."[92] He argues warfare throughout the ages has always tacitly assumed a basic equality of moral risk between combatants–kill or be killed. The equality of moral risk legitimizes violence in war through self-defense. One side's ability to kill with impunity, however, voids this

---

[88] Reuben E. Brigety II, *Ethics, Technology, and the American Way of War: Cruise Missiles and US Security Policy* (New York: Routledge, 2007), 4 – 5.
[89] Brigety, *Ethics, Technology, and the American Way of War*, 5.
[90] Brigety, *Ethics, Technology, and the American Way of War*, 134.
[91] Noel Sharkey, "Cassandra or False Prophet of Doom: AI Robots and War," *IEEE Intelligent Systems* 23, no. 4 (July-August 2008): 14 – 17.
[92] Michael Ignatieff, *Virtual War: Kosovo and Beyond* (New York: Metropolitan Books, 2000), 4.

contract, making the war unjust.[93] Ignatieff argues, "If war in the future is sold to voters with the promise of impunity they [national leaders] may be tempted to throw caution into the winds. If military action is cost-free, what democratic restraints will remain on the resort to force?"[94] Ignatieff believes virtual war has the potential to undermine the democratic peace theory by removing the risks of war to the citizens.[95]

Mike Moore echoes Ignateiff's theme. The improvement in the precision of conventional weapons leads to moral hazard.[96] For Moore, the precision US forces can achieve to avoid civilian casualties "seems like a moral plus – if, to echo Shakespeare, the cause be just and the 'quarrel honorable.'"[97] Yet, he also sees a potential for abuse of this capability.

> The United States finally possesses the technology to win conventional wars while minimizing civilian deaths…The United States can now put bombs into those metaphorical pickle barrels. That degree of accuracy has moral implications. If a war must be fought, it ought to [be] fought as cleanly as possible.

---

[93] Ignatieff, *Virtual War*, 161.
[94] Ignatieff, *Virtual War*, 179.
[95] Ignatieff, *Virtual War*, 180. For more information on the Democratic peace theory, see Edward Mansfield and Jack Snyder, *Electing to Fight: Why Emerging Democracies go to War* (Cambridge, MA: MIT, 2005).
[96] Moral hazard describes the tendency of an individual to take undue risks when the individual will not bear the costs of taking the risk. The term originated in the insurance industry, but economists co-opted the term in the 1960s while studying decision-makers under conditions of uncertainty. See Allard E. Dembe and Leslie I. Boden, "Moral Hazard: A Question of Morality?" *New Solutions: A Journal of Environmental and Occupational Health Policy* 10, no. 3 (2000): 257 – 279. Dembe and Boden lament how economists state that their use of the term moral hazard merely describes behavior and does not make a value judgment. Using this term to describe inefficient outcomes, however, makes an implicit normative judgment according to Debme and Boden. These implicit normative judgments can disparage the conduct and motives of the insured despite the framing the analysis in the dispassionate framework of economics. Debme and Boden suggest, therefore, that economics utilize value-neutral language to describe the system. When discussing moral hazard in relation to a decision to go to war, a similar normative judgment may occur. For one of the initial uses of the term moral hazard in economics, see Kenneth J. Arrow, "Uncertainty and the Welfare Economics of Medical Care," *The American Economic Review* 53, no. 5 (Dec 1963): 941 – 973.
[97] Mike Moore, *Twilight War: The Folly of U.S. Space Dominance* (Oakland, CA: The Independent Institute, 2008), 107.

> And yet, as always, the Law of Unintended Consequences comes into play. If Americans can now do war "*right*," then will they be inclined to do war more often? Do precision weapons lower the threshold for going to war? Do they inspire overconfidence–a kind of only-we-can-do-it-right hubris? If so, does that arrogance encourage national leaders to choose the war option too quickly, leaving other possibilities largely unexplored?[98]

The UN is also wary of the moral hazards created by modern technology. In a study on targeted killings, the UN notes the use of drones can create moral hazard. With targeted killings, the moral hazard emerges when decision-makers expansively interpret legal limitations on whom the State can kill and under what circumstances. As with a lethal autonomous system, the moral hazard arises due to the potential for decision-makers to discount the risk of action because the decision-maker does not have to place forces at risk during the operation. To mitigate this risk of moral hazard, the UN encourages States to ensure targeting criteria does not differ based on the type of weapon used.[99]

Asaro, however, concedes this line of argument is not limited to the development of autonomous systems. It also applies to the development of most military technologies designed to produce a technological advantage. Arkin argues, "The argument degenerates to the relinquishing of all military-related research, something that is not likely to happen."[100]

Carrying Asaro's concession further, Bradley Strawser of the University of Connecticut proffers the principle of unnecessary risk.[101] According to this concept, it is morally wrong for a commander to order soldiers to assume *unnecessary* lethal risk when carrying out a just

---

[98] Moore, *Twilight War*, 126 – 127. In this instance the moral hazard is described in moralistic, rather than value-neutral, terms and indicates an implicit normative judgment.
[99] United Nations General Assembly, *Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, Philip Alston*, Addendum, *Study on Targeted Killings*, A/HRC/14/24/Add.6, 28 May 2010, 25 (para. 80).
[100] Arkin, *Governing Lethal Behavior in Autonomous Robots*, 41.
[101] Bradley J. Strawser, "Moral Predators: The Duty to Employ Uninhabited Aerial Vehicles," *Journal of Military Ethics* 9, no. 4 (December 2010): 342 – 368.

action for a good end.[102] Strawser asserts, "If a given action can be equally well accomplished via two different methods, one of which incurs less risk for the warfighter's personal safety than the other, then a justification must be given for why this safer method is not used."[103] Based on the principle of unnecessary risk, Strawser argues the use of remotely piloted aircraft, such as the General Atomics MQ-9 Reaper, to conduct lethal attacks is a moral imperative assuming their use is in pursuit of a "morally justified yet inherently risky action."[104]

American University Law Professor Kenneth Anderson also refutes the argument that the resort to force can become too easy with technology.[105] For Anderson, the conceptual approach of this argument is wrong. Based on this line of reasoning, he argues, a successful strategy becomes "immoral, not because of the damage it causes achieving its success, but because success itself increases the propensity to do it too much."[106] When viewed through the lens of economics, the ease of warfare argument seemingly implies the existence of efficient or inefficient levels of disincentives for the use of force.[107] Consequently, the

---

[102] Strawser, "Moral Predators," 344. Strawser argues the principle of unnecessary risk justifies the use of armed remotely piloted vehicles. He does not believe, however, the principle of unnecessary risk would justify the use of autonomous lethal systems but does not expound upon this statement. He merely refers to the work of Asaro and Sparrow. See Strawser, "Moral Predators," 350.
[103] Strawser, "Moral Predators," 349.
[104] Strawser, "Moral Predators," 343.
[105] Kenneth Anderson, "Efficiency in Bello and ad Bellum: Targeted Killing Through Drone Warfare," *American University Working Paper Series*, 23 September 2011, 3.
[106] Anderson, "Efficiency in Bello and ad Bellum," 3.
[107] Anderson, "Efficiency in Bello and ad Bellum," 4. Anderson asserts the argument implies the existence of a Pareto efficiency point for the resort to force. Pareto efficient choices optimize the allocation of goods. At a Pareto inefficient point, a producer may change the allocation of goods to result in some individuals becoming better off without making anyone worse off. Conversely, the producer cannot change the allocation of products without making anyone worse off at a Pareto efficient point. The locus of Pareto efficient points for the allocation of two goods (for example, guns and butter) forms the production-possibility frontier. Anderson argues the "ability to compare of 'buy off' in Coasean bargaining to reach the efficient point in resort to force is not really there" since winning and losing in war are about different things for warring societies. Coase bargaining occurs between parties with conflicting interests and conflicting costs and benefits. If the parties are part of the same enterprise, someone from within the enterprise will make a rational choice to internalize the aggregate costs and benefits of

virtues derived from technology for *jus in bello* by lowering risk to combatants and noncombatants concomitantly create a vice for *jus ad bellum* by increasing the propensity to resort to force.[108]

Anderson argues the belligerents lack the common ground necessary to define an efficient point for the resort to force, because each side has a different notion about what conditions constitute winning and losing. Therefore, the notion of efficiency in resort to force decision "turns out to be incoherent."[109] Consider the Vietnam War. The technological advantages of the United States appeared to "foreclose any chance of victory" for North Vietnam.[110] These advantages seemingly presented an efficiency point to American policy-makers for the resort to force decision at a level unacceptable to the North Vietnamese. Yet the political objectives, and the commitment to those objectives, differed sharply for the United States and North Vietnamese. By 1972, the objective for the United Sates was to maintain a stable situation in South Vietnam.[111] Conversely, the objective for North Vietnam was to unify the nation.[112] Since winning was about different interests for the two nations, the efficiency point for the resort to force did not actually exist. Negotiations between the United States and North Vietnam therefore failed to establish a stable peace.[113]

The decision about whether to resort to the use of force is an "inherently wicked" problem requiring "elusive political judgment for

---

both parties to reach an efficiency point. When the parties are not part of the same enterprise, the market serves as the social mechanism to achieve the efficiency point. To reach the efficiency point requires a mechanism of commonality, typically money, between the two parties. See Anderson, "Efficiency in Bello and ad Bellum," 19 – 21.

[108] Anderson, "Efficiency in Bello and ad Bellum," 13.

[109] Anderson, "Efficiency in Bello and ad Bellum," 21.

[110] Stephen Randolph, *Powerful and Brutal Weapons* (Cambridge, MA: Harvard University, 2007), 2.

[111] Randolph, *Powerful and Brutal Weapons*, 9.

[112] Randolph, *Powerful and Brutal Weapons*, 23.

[113] Randolph, *Powerful and Brutal Weapons*, 333 - 340.

resolution."[114] Each side resolves this wicked problem based on what it considers winning. Lacking a common social criterion to define an efficiency point drives the normative aspect of the argument about the resort to force back to the moral question–which side is right, or which side has the just cause.[115] Therefore, Anderson concludes the resort to force is efficient when "force is resorted to justly."[116]

Arguments Against Autonomy: Technological Pessimism

In an article for *Armed Forces Journal*, Army Major Daniel Davis notes a reluctance to talk about what happens to the human role in war as armed machines gain increased intelligence, capability, and autonomy. He stakes a firm claim in the camp of retaining lethal decisions in human minds.[117] He argues that the notion of a machine appearing within the next decade able to make a lethal decision is "delusional." Davis asserts an autonomous system "doesn't have intuition…doesn't have compassion and cannot extend mercy."[118] He further argues technology simply could not substitute for the experience and intuition of a trained warfighter. To highlight this, he states, "A computer, independently running a killing machine (robot), can only carry out its programming; it will kill if ordered to do so, even in a situation in which a well-trained soldier would recognize the best thing to do is to hold fire."[119]

Human soldiers, unfortunately, do not always extend compassion and mercy on the battlefield. After interviewing soldiers who participated

---

[114] Horst W.J. Rittel and Melvin M. Webber, "Dilemmas in a General theory of Planning," *Policy Science* 4 (1973), 160. Rittel and Webber argue wicked problems are resolved, not solved. "Social problems are never solved. At best they are only re-solved–over and over again."
[115] Anderson, "Efficiency in Bello and ad Bellum," 21.
[116] Anderson, "Efficiency in Bello and ad Bellum," 21.
[117] Maj Daniel L. Davis, "Who Decides: Man or Machine?" *Armed Forces Journal*, November 2007, http://www.armedforcesjournal.com/2007/11/3036753 (accessed 16 February 2012).
[118] Davis, "Who Decides: Man or Machine?"
[119] Davis, "Who Decides: Man or Machine?"

in Operation Iraqi Freedom, the Army Surgeon General's Office Mental Health Advisory Team concluded a majority of Marines and US soldiers did not believe that they should treat Iraqi non-combatants with dignity and respect.[120] Additionally, well over a third of the soldiers and Marines believed it would be acceptable to use torture to save the life of a team member.[121] Not surprisingly, the Mental Health Advisory Team noted an increase in the mistreatment of Iraqi non-combatants by Marines and US soldiers when a unit member became a casualty.[122]

Davis, though, does capture a salient point for those developing lethal autonomous systems when he states, "robotic warfare fails to consider … the ability of potential enemy forces to leverage technology for their own purposes."[123] Here, one can detect strains of the Prussian strategist Carl von Clausewitz. For Clausewitz, war is "the collusion of two living forces … he [the enemy] dictates to me as much as I dictate to him."[124] An intelligent foe will react to and attempt to counter technological advances on the battlefield. When faced with battlefield advances, a sensible enemy will focus on developing countermeasures against the technology that "seems most dangerous at time."[125] According to Edward Luttwak's paradoxical logic of strategy, therefore, less successful technologies may retain their utility because of an

---

[120] Department of the Army, *Mental Health Advisory Team IV Operation Iraqi Freedom 05 – 07 Final Report* (Washington, DC: Office of the Surgeon General, 17 November 2006), 35, http://www.armymedicine.army.mil/reports/mhat/mhat_iv/MHAT_IV_Report_17NOV06.pdf (accessed 17 February 2012).
[121] Department of the Army, *Mental Health Advisory Team IV Operation Iraqi Freedom 05 – 07 Final Report* 35.
[122] Department of the Army, *Mental Health Advisory Team IV Operation Iraqi Freedom 05 – 07 Final Report*, 42.
[123] Davis, "Who Decides: Man or Machine?"
[124] Carl von Clausewitz, *On War* ed. and trans. Michael Howard and Peter Paret (Princeton, NJ: Princeton University, 1984), 77.
[125] Edward Luttwak, *Strategy: The Logic of War and Peace* (Cambridge, MA: Harvard University, 2001), 28 – 29.

enemy's efforts to counter, and perhaps obviate, the more successful technologies.[126]

Davis's polemic also captures an underlying uneasiness with advanced technology. The creation of technology has unintended consequences, and Finagle's Law–"Anything that can go wrong, will"–applies.[127] The etymology of the word robot captures this uneasiness. Robot evolved from the Czech word *robota*, which describes the work a landowner demanded of a peasant.[128] Two competing concepts of a robot exist: an automaton, a completely obedient machine; and an artificial person with its own intentions and capable of unexpected behavior.[129] As early as *Frankenstein*, science fiction authors have raised the specter of human creations turning on and destroying humans, which has created an underlying fear of robots, as exemplified by the *Terminator* series of movies.[130] This underlying fear is a manifestation of technological pessimism.

Massachusetts Institute of Technology (MIT) Professor Leo Marx describes technological pessimism as a "sense of disappointment, anxiety, even menace, that the idea of 'technology' arouses in many

---

[126] Luttwak, *Strategy*, 29. Luttwak avers a paradoxical logic pervades the entire realm of strategy. The paradoxical logic violates normal logic by "inducing the coming together and reversal of opposites." For a more detailed explanation of the paradoxical logic of strategy, see Luttwak, *Strategy*, 2.

[127] Bill Joy, "Why the Future Doesn't Need Us," *Wired* 8.04 (April 2000), http://www.wired.com/wired/archive/8.04/joy.html (accessed 23 February 2012), 1.

[128] Peter W. Singer, *Wired for War: The Robotics Revolution and Conflict in the 21st Century* (New York: Penguin, 2009), 66. See also, Krishnan, *Killer Robots,* 13

[129] Krishnan, *Killer Robots*, 13 and Patrick Lin, George Bekey, and Keith Abney, *Autonomous Military Robotics: Risk, Ethics, and Design*, Office of Naval Research Report N00014-07-1-1152 (San Luis Obispo, CA: California Polytechnic State University, 20 December 2008), 100.

[130] Arkin, *Governing Lethal Behavior in Autonomous Robots,* 56. Mary Shelley, *Frankenstein* (New York: SoHo Books, 2010). James Cameron, *The Terminator* (Los Angeles, CA: MGM, 1984). James Cameron, *Terminator 2: Judgment Day* (Los Angeles, CA: Carolco, 1991). Jonathan Mostow, *Terminator 3: Rise of the Machines* (Los Angeles, CA: Warner Brothers, 2003). Joseph McGinty Nichol, *Terminator Salvation* (Los Angeles, CA, Warner Brothers, 2009).

people."[131] Marx points to the destructive social and ecological consequences resulting from spectacular technological disasters, such as Bhopal, Chernobyl, the Exxon *Valdez*, acid rain, global warming, and ozone depletion as principal contributors to technological pessimism.[132] The 2003 Space Shuttle *Columbia* disaster and the 2010 *Deepwater Horizons* oil spill in the Gulf of Mexico are two contemporary examples of technological disasters contributing to the sense of technological pessimism.

Ambivalence concerning the effects of technology arises from the difficulties of divining the consequences of a particular technological innovation. Marx uses modern advances in medicine and social hygiene to illustrate this point. Modern medicine is both a blessing and a curse. In modernized countries, medicine has cured diseases, prolonged life-spans, and lowered death rates. Conversely, though, in underdeveloped countries, modern medicine has contributed to an explosive population growth with many unforeseen consequences. Technology produces both pessimism and optimism. Understanding a group's response to a technological advance requires comprehension of the group's historical context and expectations of technology's ability to be a driving force of progress.[133]

For those who oppose the technocratic idea of progress, the creation of large, complex technological systems results in an objectionable tendency to "bypass moral and political goals by treating advances in the technical means as ends in themselves."[134] Technological sophistication masks the choice of ends. If these ends, however, are "deformed, amoral, and irrational," then the "disjunction of

---

[131] Leo Marx, "The Idea of 'Technology' and Postmodern Pessimism," in *Does Technology Drive History? The Dilemma of Technological Determinism*, ed. Merritt Roe Smith and Leo Marx (Cambridge, MA: MIT, 1994), 238.
[132] Marx, "The Idea of 'Technology' and Postmodern Pessimism," 238.
[133] Marx, "The Idea of 'Technology' and Postmodern Pessimism," 238 – 239.
[134] Marx, "The Idea of 'Technology' and Postmodern Pessimism," 254.

means and ends becomes particularly risky."[135] Infatuation with the technological challenge in developing something as complex as a lethal autonomous system has the potential to mask the underlying reasons for the system's development.

Arguments Against Autonomy: Military Ethos

Air Force Lieutenant Colonel Michael Contratto argues the use of lethal autonomous systems will have a deleterious effect on the military ethos. These systems would chip away at the moral foundation of the profession of arms by relinquishing "direct moral agency for war's most profound activities."[136] He characterizes the development of these systems as an outsourcing of the traditional warrior role, which will expedite a decline in the military profession.[137] It also portends a decline in the military ethic. Exposure to physical risk and making profound wartime decisions requires the exercise of physical and moral courage. This physical and moral courage forged in the crucible of combat creates a key ingredient in the nation's moral stock, which would diminish with the use of these systems.[138] Autonomous lethal engagements would also erode professional ethics that require the delineation of moral authority for actions within the profession. Professional ethical codes enable the members of a profession to undertake actions typically impermissible in moral terms yet necessary for the operation of society. In the conduct of their duties, professional codes enable doctors to prescribe addictive and dangerous drugs to patients and lawyers to conceal facts concerning the crimes of clients. Contratto argues the military should be "no more eager to give up this weighty responsibility to an autonomous agent than

[135] Marx, "The Idea of 'Technology' and Postmodern Pessimism," 254 – 255.
[136] Lt Col Michael Contratto, "The Decline of the Military Ethos and Profession of Arms: An Argument Against Autonomous Lethal Engagements," (master's thesis, Air War College, 2011), 15.
[137] Contratto, "The Decline of the Military Ethos and Profession of Arms," 16.
[138] Contratto, "The Decline of the Military Ethos and Profession of Arms," 16 – 18.

doctors would forgo their responsibility to operate, lawyers to defend, or judges to adjudicate."[139]

Critiques of autonomy along the lines of Davis and Contratto hint at an underlying argument against artificial intelligence, a critical component of autonomous systems. Applying artificial intelligence in the military environment, particularly for command decisions, would undermine the human element in war. Leadership is a key aspect of command stemming from the human element. A commander's presence at the front builds confidence and improves morale among the soldiers.[140] Brigadier General Huba Wass de Czege argues autonomous systems can "predict factors in war that are based on the laws of physics, but they are unreliable predictors of moral factors–the human element."[141] Managing people and practicing the art of leadership are critical elements of command, which if missing, would reduce the efficacy of artificial intelligence systems.[142]

The critiques of Davis and Contratto also allude to a criticism of postheroic warfare. With the exceptions of self-defense and national survival, postheroic warfare tolerates combat "by remote bombardment alone, without soldiers at risk on the ground."[143] Casualty aversion necessitates the methods of postheroic warfare and emerges in postindustrial societies from prosperity and demographic changes.[144] The decreasing family sizes in postindustrial societies has confined death, for the most part, to the very old, making exposure to death itself

---

[139] Contratto, "The Decline of the Military Ethos and Profession of Arms," 18 – 19.
[140] David Lonsdale, *The Nature of War in the Information Age: Clausewitzian Future* (New York: Frank Cass, 2004), 120 – 121.
[141] Brigadier General Huba Wass de Czege quoted in Lonsdale, *The Nature of War in the Information Age*, 121.
[142] Lonsdale, *The Nature of War in the Information Age*, 120 – 121.
[143] Luttwak, *Strategy*, 74. For an additional perspective on postheroic warfare, see Edward N. Luttwak, "Toward Post-Heroic Warfare," *Foreign Affairs* 74, no. 3 (May/June 1995), 109 – 122.
[144] Luttwak, *Strategy*, 68 – 72.

a less common occurrence in the human experience.[145] While the loss of a child has always been a tragic event, declining family sizes has made the loss of a young family member, particularly from combat, an "extraordinary and fundamentally unacceptable event."[146] Determined political leaders can widen their freedom of action with exceptional effort, as illustrated by the actions of President George H.W. Bush in the Gulf War and Prime Minister Margaret Thatcher in the Falkland Islands.[147] The Great Power responsibility of restoring order, however, remains, and leaders are not always able or willing to expend the necessary political capital to expand their freedom of action.[148] Precision aerial bombing, therefore, provides a partial solution.[149]

Paul Kahn of Yale Law School argues that the technological capabilities enabling postheroic warfare have given risk to the paradox of riskless war. In war, a combatant may legally and morally injure or kill an enemy combatant, because each is acting in self-defense vis-à-vis the other.[150] Militaries, however, have an ethical obligation to minimize the risk its own soldiers face. To accomplish this, militaries strive to create an asymmetric condition by which the "enemy suffers the risk of injury while its own forces remain safe."[151] The paradox of riskless warfare emerges when the asymmetry of capabilities undermines the reciprocal imposition of risk on combatants. Kahn argues riskless warfare stresses the "limits of the traditional moral justification of combat."[152] He concedes that an application of force with a dramatic asymmetry "might be morally justified" and "might be used to promote morally appropriate

[145] Luttwak, *Strategy*, 71.
[146] Luttwak, *Strategy*, 71.
[147] Luttwak, *Strategy*, 73.
[148] Luttwak, *Strategy*, 74.
[149] Luttwak, *Strategy*, 74.
[150] Paul W. Kahn, "The Paradox of Riskless Warfare," *Yale Law School Faculty Scholarship Series* (2002), Paper 326, 2.
[151] Kahn, "The Paradox of Riskless Warfare," 2.
[152] Kahn, "The Paradox of Riskless Warfare," 2.

ends." [153] Yet, a moral conundrum develops, because, "Without the imposition of mutual risk, warfare is not war at all."[154] The lack of an imposition of mutual risk creates a moral dilemma arising from the notion of the moral equality of soldiers.[155] War establishes an "equal right to kill" for soldiers.[156] Society distinguishes war from murder by the restrictions placed "on the reach of battle."[157] The solution to the problem, according to Kahn, lies in changing the application of force paradigm from warfare to policing, whereby the application of force is against the morally guilty.[158]

During operations in Kosovo, the North Atlantic Treaty Organization (NATO) imposed altitude restrictions on its pilots to decrease their risk exposure.[159] Kahn points to these operations as lacking reciprocal risk for NATO pilots.[160] Kahn's paradox of riskless war echoes Ignatieff's criticism of NATO's Kosovo operations as virtual war. Though written before the increased use of drones to conduct lethal strikes in Iraq and Afghanistan, Kahn's paradox of riskless war adumbrates criticism of drone strikes as diminishing the moral equality of combatants through the removal of risk for one side. This line of reasoning led Strawser to defend the use of drones in Afghanistan and Iraq through the principle of unnecessary risk.[161]

While many compelling arguments against autonomy exist, equally compelling arguments for autonomy also exist, which is where attention will now turn.

---

[153] Kahn, "The Paradox of Riskless Warfare," 3.
[154] Kahn, "The Paradox of Riskless Warfare," 4.
[155] For more on the notion of the moral equality of soldiers, see Walzer, *Just and Unjust Wars*, 34 – 47.
[156] Walzer, 41.
[157] Walzer, 42.
[158] Kahn, "The Paradox of Riskless Warfare," 7.
[159] Ingatieff, *Virtual War*, 62.
[160] Kahn, "The Paradox of Riskless Warfare," 4
[161] Strawser, "Moral Predators," 343.

Arguments for Autonomy

Applying technology to solve battlefield problems is seductively alluring for the US military. Despite a separation of nearly thirty years, Army Chief of Staff General William Westmoreland and Air Force Chief of Staff General Ronald Fogelman had strikingly similar visions for the application of technology on the battlefield.[162] In 1969, General Westmoreland argued, "On the battlefield of the future, enemy forces will be located, tracked, and targeted almost instantaneously through the use of data links, computer assisted intelligence evaluation, and automated fire control."[163] General Fogelman echoed this theme when he asserted, "In the first quarter of the 21st century you will be able to find, fix or track, and target–in near real-time–anything of consequence that moves upon or is located on the face of the Earth."[164]

Fulfilling this vision will require copious amounts of data. The shear amount of data generated and the processing power necessary to sift through this data has the potential to overwhelm human decision-makers. This trend has been building for many years. In 1961, John Kemeny of RAND argued, "Modern war has become too complex to be entrusted to the intuition of even the most experienced military commander. Only our giant brains [computers] can calculate all the possibilities."[165] The critical task becomes determining whether, and how, to integrate artificial intelligence and autonomy while bearing in mind war witnesses the confluence of policy, humanity, uncertainty, friction, and intelligent foes in a complex and dynamic undertaking.[166]

---

[162] Antoine Bousquet, *The Scientific Way of Warfare: Order and Chaos on the Battlefields of Modernity* (New York: Columbia, 2009), 217.
[163] General William Westmoreland's address to the Association of the US Army, October 14, 1969, quoted in Bousquet, *The Scientific Way of Warfare*, 126.
[164] General Fogelman quoted in Antoine Bousquet, *The Scientific Way of Warfare*, 217.
[165] John Kemeny quoted in Bousquet, *The Scientific Way of Warfare*, 121.
[166] Lonsdale, *The Nature of War in the Information Age*), 112.

Arguments for Autonomy: The Digital Imperative

Computers can clearly process certain forms of information faster than humans can. The *USAF UAS Flight Plan* foresees unmanned systems executing actions at speeds faster than the human decision making cycle. When this occurs, the human operators will "no longer be 'in the loop' but rather 'on the loop' – monitoring the execution of certain decisions."[167]

Sun Tzu, the ancient Chinese master of warfare, advises, "Speed is the essence of war."[168] The use of artificial intelligence and autonomous systems confers an advantage of speed in decision-making. When thinking about speed in decision-making, many turn to John Boyd's Observe-Orient-Decide-Act (OODA) loop. The OODA loop provides a cognitive theory for decision-making. In the observe stage, the decision-maker absorbs information from the environment. The decision-maker then interprets the information through frameworks of analysis during the orientation stage. During the decide stage, the decision-maker commits to a particular course of action to implement during the act stage.[169]

Warfare becomes a competition between opposing OODA loops in which the more effective OODA loop prevails. Victory results from completing the OODA loop cycle faster than the adversary or from changing operational tempo and rhythm such that the adversary cannot keep up.[170] Victory may also emerge by disrupting or corrupting the adversary's OODA loop.

---

[167] Department of the Air Force, *United States Air Force Unmanned Aircraft Systems Flight Plan 2009 – 2047* (Washington, DC: Headquarters Air Force, 18 May 2009), 41.
[168] Sun Tzu, *The Illustrated Art of War*, trans. Samuel Griffith (New York: Oxford University, 2005), 213.
[169] Bousquet, *The Scientific Way of Warfare*, 188.
[170] Bousquet, *The Scientific Way of Warfare*, 194. For an analysis on how the capabilities of the Air Operations Center have enabled a Joint Forces Air Component Commander to shorten the OODA loop, see Michael W. Kometer, *Command in War: Centralized Versus Decentralized Control of Combat Airpower* (Maxwell AFB, AL: Air University, 2007), 185 – 212.

Creating a quicker OODA loop than the adversary, however, is not just a matter of cycling through the sequence faster. This simplistic reading of Boyd's theory misses the feedback mechanisms and cross connections embedded within the OODA loop which make it a complex adaptive system. Changing tempo and rhythm requires initiative, surprise, and deception. As Antoine Bousquet of the University of London argues, "Merely increasing the speed at which one acts by responding to stimulus from pre-established templates (i.e. without truly orienting) is not a quickening of the OODA 'loop,' a point missed by many subsequent theorists."[171] While speed in decision-making is an important aspect of Boyd's OODA loop, the aim of Boyd's concept is to "render the enemy powerless by denying him the time to mentally cope with the rapidly unfolding, and naturally uncertain, circumstances of war, and only in the most simplified way, or at the tactical level, can this be equated with the narrow, rapid OODA loop idea."[172] This more nuanced understanding of the OODA loop highlights the importance of not only speed but also tempo in the decision-making cycle.[173]

Clausewitz presages this nuanced understanding of the OODA loop with the concept of *coup d'oeil*–an indispensible quality of the military genius. *Coup d'oeil* encapsulates the ability of the military genius to make not only rapid and accurate decisions but also the ability to achieve a "quick recognition of a truth that the mind would ordinarily miss or would perceive only after long study and reflection."[174] Advancements in digital and networked technologies, such as those provided by Blue Force Tracking capabilities, can augment the commander's *coup d'oeil* and facilitate more effective decision-making. For example, during Operation IRAQI FREEDOM (OIF), Blue and Red

---

[171] Bousquet, *The Scientific Way of Warfare*, 195.
[172] Frans Osinga, *Science, Strategy and War: The Strategic Theory of John Boyd* (New York: Routledge, 2007), 237.
[173] Lonsdale, *The Nature of War in the Information Age*, 114.
[174] Clausewitz, *On War*, 102.

Force Tracking provided US Central Command (CENTCOM) Commander General Tommy Franks and his staff with an "unprecedented level of awareness."[175] The confidence from this awareness enabled General Franks to direct coalition forces to secure Iraqi oil fields ahead of schedule during the opening hours of the 2003 invasion of Iraq because he knew the Iraqi military was not in a position to respond.[176]

Speed and tempo in the decision-making cycle can have decisive battlefield effects. In OIF, General Franks utilized the speed and maneuver capabilities of coalition forces to "disrupt the Iraqi's ability to react effectively" and "[get] inside the enemy's decision cycle."[177] Conversely, slowness or an inability to control the tempo of the decision-making process can lead to disastrous consequences. For instance, the failures during World War I to achieve breakouts from trench warfare resulted in part from the command and control tempo being insufficient to exploit breakthroughs.[178]

A danger exists that a nation not willing to allowing computers to process information and to make certain decisions could fall behind an adversary willing to do so. David Lonsdale of Kings College London terms this the digital imperative–the "pressure to employ AI [artificial intelligence] in command for fear that the enemy may do so whilst you do not."[179] The digital imperative is a security dilemma performed on a cyber-age stage.

Lonsdale argues artificial intelligence will assume a greater role in making command decisions for two primary reasons: (1) a computer's computational and data storage capabilities and (2) human weaknesses. Lonsdale sees several advantages stemming from a computer's computation and data storage capabilities. A digitized force is able to

---

[175] Tommy Franks, *American Solider* (New York: HarperCollins, 2004), 448.
[176] Franks, *American Solider*, 448.
[177] Franks, *American Solider*, 466.
[178] Lonsdale, *The Nature of War in the Information Age*, 114.
[179] Lonsdale, *The Nature of War in the Information Age*, 115.

increase its operational tempo. A computer's shear computational capability could help a commander develop the necessary campaign options and contingency plans for the operation. A computer could store vast amounts of information on previous campaigns and commanders to provide a detailed historical analysis. Additionally, artificial intelligence may be necessary just to cope with the avalanche of data generated by modern warfare.[180]

Lonsdale also sees an impetus for utilizing artificial intelligence due to human weaknesses. A computer is immune to emotions and psychological pressures.[181] Computers do not suffer from human physical and mental limitations–the machine does not become tired, hungry, or upset.[182] Finally, artificial intelligence will have the "moral courage required to bear the responsibility of command."[183]

As an example of the moral courage required by a commander, Lonsdale cites Air Marshall Arthur Harris's decision to approve the Millennium raid against Cologne in May 1942. In approving this raid, Air Marshall Harris staked virtually his entire force on one raid at a time when German air defenses were inflicting significant losses on large Bomber Command raids. The official history describes the decision facing Air Marshall Harris as having the potential to be either a "great triumph" or "irremediable."[184] Such a grave decision required moral courage to bear the responsibilities of the decision.

Another example of a wartime decision requiring moral courage stems from World War II submarine operations in the Pacific Ocean. The US Navy changed its concept of operations for submarine warfare from attacking the Japanese fleet to raiding merchant shipping.[185] Prewar

---

[180] Lonsdale, *The Nature of War in the Information Age*, 114 – 116.
[181] Lonsdale, *The Nature of War in the Information Age*, 115.
[182] Lonsdale, *The Nature of War in the Information Age*, 117.
[183] Lonsdale, *The Nature of War in the Information Age*, 117.
[184] Lonsdale, *The Nature of War in the Information Age*, 118.
[185] Stephen P. Rosen, *Winning the Next War: Innovation and the Modern Military* (Ithaca, NY: Cornel University, 1991), 130 – 131.

training for submarines built "extreme caution" into the concept of operations.[186] Wartime success against Japanese merchant vessels, however, required innovations in tactics considered "wildly dangerous by prewar standards."[187] Submarine skippers required moral courage to innovate and develop tactics contrary to prewar training. The Navy relieved 30 percent of the submarine commanders for cause because these men were unable make this required transition.[188]

Lonsdale sees artificial intelligence as insurance against the potential limitations in a military commander not as well endowed with moral courage as Air Marshall Harris. Such commanders may hesitate to act decisively or react to battlefield changes without the necessary flexibility. For example, General McClellan's reticence to act decisively led an exasperated President Lincoln to lament, "If General McClellan does not want to use the Army for some days, I would like to borrow it provided I could see how it could be made to do something."[189] General McClellan's "ruinous levels of undue caution" ultimately led to his dismissal as commander of the Army of the Potomac.[190] A command system augmented with artificial intelligence could help avoid these human shortfalls. Lonsdale further argues artificial intelligence would also avoid the human peccadillo of overconfidence as exemplified by commanders such as Napoleon or Hitler.[191]

As Lonsdale implies, human decision-making is problematic. Split-second decisions, those made in less than 500 milliseconds, are more prone to errors than deliberate decisions. Humans make split-second

---

[186] Rosen, *Winning the Next War*, 135.
[187] Rosen, *Winning the Next War*, 138.
[188] Rosen, *Winning the Next War*, 131.
[189] Quoted in Jeffrey M. Reilly, *Design: Distilling Clarity for Decisive Action* (Maxwell Air Force Base, AL: Air Command and Staff College [Department of Joint Warfare Studies], October 2011), 5. For a detailed discussion of General McClellan's indecisiveness during the 1862 Peninsula Campaign and Second Battle of Manassas of the American Civil War, see Reilly, *Design*, 4 – 8.
[190] Lonsdale, *The Nature of War in the Information Age*, 118.
[191] Lonsdale, *The Nature of War in the Information Age*, 118.

decisions based not on underlying evidence but rather on stereotypes and preconceived notions. Better decisions emerge when humans use a combination of deliberate and instinctive thinking and by reducing complex problems to their simplest elements.[192]

Unfortunately, even the human deliberative thinking process does not always work well, particularly when dealing with complex systems. Dietrich Dörner studied the human decision-making process involving complex systems. In one case study, he examined the decisions leading to the 1986 meltdown of the Chernobyl nuclear reactor. Dörner concluded that "the slowness of our [human] thinking and the small amount of information we can process at any one time, our tendency to protect our sense of our competence, the limited inflow capacity of our memory, and our tendency to focus only on immediate pressing problems … cause the mistakes we make in dealing with complex systems."[193] Computers and artificial intelligence offer the prospect of mitigating these weaknesses in human decision-making.

In addition to having psychological and physiological weaknesses, humans are expensive. When the Navy examined the total ownership cost of a *Nimitz*-class carrier, it found the personnel costs were "the long pole in the cost tent" over the ship's lifecycle.[194] Personnel expenses, therefore, have the potential to create an incentive for automation to reduce costs.[195]

---

[192] Malcolm Gladwell, *Blink: The Power of Thinking Without Thinking*, (New York: Little, Brown and Company, 2005), 141, 212.
[193] Dietrich Döner, *The Logic of Failure – Why Things Go Wrong and What We Can Do to Make Them Right*, trans. Rita and Robert Kimber, (New York: Metropolitan Books, 1996), 193.
[194] Paul S. Canning, *A Definitive Work on Factors Impacting the Arming of Unmanned Vehicles*, NSWCDD TR-05/36 (Dahlgren, VA: Department of the Navy, May 2005), 13.
[195] Canning, *A Definitive Work on Factors Impacting the Arming of Unmanned Vehicles*, 13.

Arguments for Autonomy: Improved Battlefield Performance

Arkin avers future autonomous systems will perform better on the battlefield than humans will.[196] An autonomous system will presumably not have the same self-preservation instinct as a human, enabling the autonomous system to act more conservatively.[197] Unlike a human in cases of uncertain target identification, an autonomous system can afford to wait for the target to shoot first. Furthermore, Arkin believes the broad range of sensors available to an autonomous system will make it better equipped to observe the battlefield than humans can. He also asserts autonomous systems will not suffer from scenario fulfillment. Due to this phenomenon, which is a form of premature cognitive closure, humans neglect contradictory information and interpret information to fit pre-existing belief patterns.[198] Arkin echoes Lonsdale's argument that autonomous systems do not suffer from emotions clouding their battlefield judgment and are therefore able to integrate information faster than humans. A unique argument Arkin offers in favor of utilizing autonomous systems is the ability of these systems to monitor the ethical behavior of human soldiers.[199] He sees these systems as means to avoid the ethical failings of human soldiers addressed in the Army's Mental Health Advisory Team Report. Arkin also sees autonomous systems providing "potential operational benefits to the military: faster, cheaper, better mission accomplishment; longer range, greater persistence, longer endurance, higher precision; faster engagement; and immunity to chemical and biological weapons among others."[200]

---

[196] Arkin, *Governing Lethal Behavior in Autonomous Robots*, 29 – 30.
[197] Daniel Brunstetter and Megan Braun argue the use of remotely piloted aircraft (RPA) will enable more proportionality and discrimination on the battlefield. See Daniel Brunstetter and Megan Braun, "The Implications of Drones on the Just War Tradition," *Ethics & International Affairs* 25, no. 3 (Fall 2011), 348 – 352.
[198] Arkin asserts scenario fulfillment may have contributed to the downing of Iran Air Flight 655 by the *USS Vincennes* in July 1988.
[199] Arkin, *Governing Lethal Behavior in Autonomous Robots*, 29 – 30.
[200] Arkin *Governing Lethal Behavior in Autonomous Robots*, 30.

The operational context necessitating increased levels of autonomy arises from operations in a contested environment.[201] The new Air-Sea Battle concept envisions operations to maintain "freedom of action in the global commons" despite an anti-access and area-denial threat.[202] Threats from advanced surface to air missiles, such as the SA-10 and SA-20, combined with emerging air-to-air threats from fifth-generation aircraft such as the Chinese J-20 and the proposed Russian-Indian PAK-FA pose a growing challenge to US air dominance.[203] The current generation of remotely piloted aircraft (RPA), however, would "struggle in enemy-controlled airspace due to a lack of survivability and insufficient capacity to respond to contingencies such as incoming threats and changes in the weather."[204] In highly contested airspace, an unmanned system may not have the luxury of waiting for human inputs in reaction to a developing threat. Therefore, some degree of autonomy becomes essential for the RPA to survive in the highly contested airspace envisioned with Air-Sea Battle.[205]

Conclusion

As Antonio reminds Sebastian in Shakespeare's *The Tempest*, the past is prologue.[206] Between the World Wars, America's airmen found themselves developing airpower theory lacking clear legal guidance and with unresolved moral and ethical questions. Today's airmen find themselves in an analogous situation regarding drone strikes and lethal autonomous systems.

---

[201] Caitlin H. Lee, "Embracing Autonomy: The Key to Developing a New Generation of Remotely Piloted Aircraft for Operations in Contested Air Environments," *Air and Space Power Journal* 24, no. 4 (Winter 2011), 78.
[202] Department of Defense, "Background Briefing on Air-Sea Battle," 9 November 2011, www.defense.gov/transcript.aspx?transcriptid=4932 (accessed 16 February 2012).
[203] Lee, "Embracing Autonomy," 78 – 79.
[204] Lee, "Embracing Autonomy," 78.
[205] Lee, "Embracing Autonomy," 82.
[206] William Shakespeare, *The Tempest*, eds. Virginia M. Vaughan and Alden T. Vaughan (London: The Arden Shakespeare, 1999), 2.1.253. References are to act, scene, and line.

While the law of armed conflict is silent on the legality of lethal autonomous systems, it provides the general principles of necessity, proportionality, discrimination, and humaneness that must guide the development of these systems. The legality of lethal autonomous systems will rest with the implementation of these principles.

Moral and ethical arguments both for and against these systems exist, yet the issue remains undecided. Critics of autonomous systems point to lingering concerns. Difficulties in enabling a machine to apply the principles of discrimination and proportionality exist. These difficulties present continuing technical challenges. Questions remain regarding legal and moral accountability for actions by an autonomous system. As the next chapter will illustrate, system design may mitigate these accountability concerns. Technological pessimists point to Finagle's Law. Again, the following chapter will demonstrate how verification and validation processes could alleviate these criticisms. Some argue these systems create conditions of moral hazard by making the resort to force too easy. Others point to a decline in the warrior ethos. Both of these arguments are inherent in most military related technological advancements. Advocates point to the digital imperative and the potential of improved battlefield performance.

Despite these lingering questions, the development of lethal autonomous systems continues. Scientists and engineers are exploring methods to develop autonomous systems capable of adhering to the law of armed conflict and of ethical reasoning. Since the legality of autonomous systems will hinge on architectural implementations enabling legal and ethical reasoning, attention now turns to exploring methods for developing an ethical autonomous system.

Chapter 2

Developing an Ethical System

"Would you tell me, please, which way I ought to walk from here?" asked Alice
"That depends a good deal on where you want to get to," said the Cat.
"I don't much care where…," said Alice.
"Then it doesn't much matter which way you walk," said the Cat.
"…so long as I get somewhere," Alice added as an explanation.
"Oh, you're sure to do that," said the Cat, "if you only walk long enough."

—Lewis Carroll
Alice in Wonderland

The Cat reminds Alice the path she ought to take depends on her intended destination. With lingering questions about her strategic guidance, the Cat reminded her, she will still arrive at a destination by beginning her journey.[1] In much the same way, persistent questions during the interregnum between the World Wars clouded the path air planners ought to take to develop doctrine and tactics for aerial bombardment in an evolving technological Wonderland. Yet, despite unresolved concerns regarding the legality and morality of conducting aerial bombardment, particularly against civilian populations, the cadre of air warfare strategists at the Air Corps Tactical School (ACTS) forged ahead to develop plans and tactics for aerial bombardment campaigns. In developing these plans, the ACTS strategists implicitly understood war is "not merely an act of policy but a true political instrument, a continuation of political intercourse, carried on with other means."[2] The

[1] Lewis Carroll, *Alice in Wonderland* (Scituate, MA: Digital Scanning, 2007), 89 – 90.
[2] Carl von Clausewitz, *On War* ed. and trans. Michael Howard and Peter Paret (Princeton, NJ: Princeton University, 1984), 87.

1931 ACTS text, *Air Force*, notes the inherent political nature of aerial bombardment campaigns. The text warns planners that the United States would conduct an aerial bombardment campaign against political objects, such as cities, when "specified by only the highest authority."[3] Furthermore, the text argues such an attack "should never be adopted except as the result of a careful estimate of the results to be accomplished."[4]

By 1938, ACTS thinking on aerial bombardment had coalesced into the industrial fabric theory. According to this theory, attacking key industries could produce decisive effects.

> The economic structure of a modern highly industrialized nation is characterized by the great degree of interdependence of its various elements. Certain of these elements are vital to the continued functioning of the modern nation. If one of these vital elements is destroyed the whole of the economic machine ceases to function. Some of these vital elements are composed of relatively few targets which are vulnerable to the air offensive. Consequently, the air offensive varies in effectiveness directly with the degree of concentration of enemy vital elements into such vulnerable targets. With maximum concentration and interdependence the air offensive may attain maximum effectiveness. Against a modern highly industrialized nation air force action has the possibility for such far reaching effectiveness that such action may produce immediate and decisive results.[5]

The industrial fabric provided a path to "undermine the enemy's national morale without raising the kinds of ethical questions that might be posed by attacks on targets considered to be 'social' or 'political' in nature."[6] Under the leadership of Lt Col Harold L. George, the Army Air Corps Air War Plans Division (AWPD) developed the comprehensive air plan, known as AWPD-1, designed to defeat the Axis power. The industrial fabric

---

[3] ACTS, *Air Force* text 1931, AFHRC file 248.101-16, 53.
[4] ACTS, *Air Force* text 1931, AFHRC file 248.101-16, 53.
[5] ACTS, "Air Offensive Characteristics" section 14 in *Air Force Air Warfare* (Feb 1, 1938), AFHRC file 248.101-1.
[6] Tami Biddle, *Rhetoric and Reality: The Evolution of British and American Ideas About Strategic Bombing, 1914 – 1945*, (Princeton, NJ: Princeton University, 2002), 160.

theory provided the foundation for AWPD-1 yet ignored the difficult questions of modern warfare.[7]

ACTS developed the industrial fabric theory for strategic bombing without the resolution of legal, ethical, and moral issues associated with the ability to expand the reach of violence. In keeping with the observation attributed to Mark Twain that "History doesn't repeat itself, but it does rhyme," a similar pattern is at risk of reoccurring today. The *USAF UAS Flight Plan* states, "Authorizing a machine to make lethal combat decisions is contingent upon political and military leaders resolving legal and ethical questions."[8] The UK Ministry of Defense echoes this theme and argues, despite exploration by academics, the moral and ethical questions associated with autonomous systems are starting "to require real-world answers."[9] Despite the uncertainties, scientists and engineers are toiling away in military and academic laboratories to develop autonomous systems capable of employing lethal force. These scientists are working to develop what they consider ethical systems by utilizing the law of armed conflict principles as a guide. Thus, Alice still does not know where she walks, but the terrain seems distantly familiar. The methods the scientists are using to develop lethal autonomous systems provide the next topic for exploration.

The Science Fiction of Robot Ethics

Science fiction often provides inspiration for technological developments. Jules Verne's 1865 work *From the Earth to the Moon* provided an inspiration for space flight.[10] In his 1907 novel *The War in the Air*, H.G. Wells created an apocalyptic vision of aerial warfare,

---

[7] Biddle, *Rhetoric and Reality*, 206 – 207.
[8] Department of the Air Force, *United States Air Force Unmanned Aircraft Systems Flight Plan 2009 – 2047* (Washington, DC: Headquarters Air Force, 18 May 2009), 41.
[9] Ministry of Defense, Joint Doctrine Note 2/11 *The UK Approach to Unmanned Aircraft Systems* (London: Office of the Assistant Head Air and Space (Development, Concepts and Doctrine), 30 March 2011), 5-8.
[10] Jules Verne, *From the Earth to the Moon: Direct in Ninety-Seven Hours and Twenty Minutes: and a Trip Round It* (New York: Charles Scribner's Sons, 1890).

provoking consideration of the promise and peril of this new form of warfare.[11] One of the inspirations for the development of an ethical system for robots stems from Isaac Asimov's Three Laws of Robotics. Asimov introduced the Three Laws in his 1941 short story "Runaround" and expanded on the Laws in the anthology *I, Robot.*[12]

Asimov's Three Laws of Robotics create an elegant, yet simple, formulation of a robot ethic.[13] The simplicity of the Laws offers the alluring promise of emergent behavior, whereby a few simple rules give rise to the enormously complex yet self-organized behavior observed in the natural world.[14] The First Law states that a robot may not injure a human being, or, through inaction allow a human to come to harm.[15] A small set of robots received a modified version of the First Law that deleted the inaction clause and said only that a robot may not harm a human.[16] According to the Second Law, a robot must obey the orders given it by humans except where such orders would conflict with the First Law.[17] The Third Law stipulates that a robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.[18] In subsequent works, Asimov adds a Zeroth Law by which a robot may not harm all humanity, or through inaction, allow humanity to come to harm.[19]

The elegant simplicity of the Three Laws belies the difficulties experienced in their implementation. As the robot developers Mike

---

[11] H.G. Wells, *The War in the Air* (New York: MacMillan, 1907)
[12] Isaac Asimov, "Runaround," *Astounding Science Fiction*, March 1942, 94 – 103, and Isaac Asimov, *I, Robot* (Garden City, NY: Doubleday, 1950).
[13] Ronald C. Arkin, *Governing Lethal Behavior in Autonomous Robots* (New York: CRC Press, 2009), 48.
[14] M. Mitchell Waldrop, *Complexity: The Emerging Science at the Edge of Order and Chaos* (New York: Simon & Schuster, 1992), 152.
[15] Asimov, *I, Robot*, 51.
[16] Asimov, *I, Robot*, 123.
[17] Asimov, *I, Robot*, 51.
[18] Asimov, *I, Robot*, 51.
[19] Isaac Asimov, *Robots and Empire* (London: Grafton, 1985) quoted in Roger Clarke, "Asimov's Laws of Robotics: Implications for Information Technology – Part II," *IEEE Computer* 27, no. 1 (Jan 1994), 57 – 66. Clarke's article is also available at http://www.rogerclarke.com/SOS/Asimov.html (accessed 23 February 2012).

Donnovan and Greg Powell discover, something "invariably" goes wrong with the robots.[20] For example, in "Runaround," the robot Speedy cannot resolve the conflict between the Second and Third Laws when trying to comply with Donnovan's direction on the planet Mercury to gather selenium near a source of carbon monoxide, which would destroy Speedy. The robots resolve conflicts among the Three Laws based on potentials assigned to each Law. In this instance, the potential for the Third Law increased because of Speedy's complexity and expense, while the potential assigned to the Second Law decreased because Donnovan did not add a sense of urgency to his command. Speedy could not proceed beyond the point at which the potentials assigned to the Second and Third Laws equalized, causing the robot to runaround in a circle at the locus of equilibrium points.[21] As "Runaround" illustrates, a rule-based morality such as the Three Laws has inherent limitations. Asimov uses the Three Laws as a literary device to explore these limitations, such as what to do when two rules conflict, whether the rules are constraints or guidelines, and which rules apply in a particular situation.[22] Thus, despite their potential to create emergent behavior, the Three Laws fall short.

Other authors have subsequently modified Asimov's Three Laws in attempts to clarify ambiguities and to resolve loopholes.[23] Roger Clarke provides an extended set of the Laws of Robotics.

> The Meta-Law: A robot may not act unless its actions are subject to the Laws of Robotics.

---

[20] Asimov, *I, Robot*, 84.
[21] Asimov, *I, Robot*, 49 – 54. Volcanic action on Mercury produced gas containing sulphur dioxide, carbon dioxide, and carbon monoxide. The carbon monoxide would combine with the robot's iron to form iron carbonyl, which damages the robot. To solve the dilemma, Powel ends putting himself at risk to increase the First Law potential.
[22] Patrick Lin, George Bekey, and Keith Abney, *Autonomous Military Robotics: Risk, Ethics, and Design*, Office of Naval Research Report N00014-07-1-1152 (San Luis Obispo, CA: California Polytechnic State University, 20 December 2008), 31.
[23] Lin, *Autonomous Military Robotics*, 31.

> Law Zero: A robot may not injure humanity, or, through inaction, allow humanity to come to harm.
>
> Law One: A robot may not injure a human being, or through inaction, allow a human being to come to harm, unless this would violate a higher-order Law.
>
> Law Two: A robot must obey orders give it by human beings, except where such orders would conflict with a higher-order Law. A robot must obey orders given it by superordinate robots, except where such orders would conflict with a higher-order Law.
>
> Law Three: A robot must protect the existence of a superordinate robot, as long as such protection does not conflict with a higher order Law. A robot must protect its own existence as long as such protection does not conflict with a higher-order Law.
>
> Law Four: A robot must perform the duties for which is has been programmed, except where that would conflict with a higher order Law.
>
> The Procreation Law: A robot may not take part in the design or manufacture of a robot unless the new robot's actions are subject to the Laws of Robotics.[24]

Despite proposing this extended set of laws, Clarke argues modifications lose the simplicity of Asimov's original laws in "complexity, legalisms, and semantic richness."[25] For Clarke, developing these laws serves a useful purpose by highlighting important issues for consideration, such as the recognition of stakeholder rights, closed-system versus open-system thinking, blind acceptance of technological imperatives, and human acceptance or rejection of robots.[26]

While science fiction may inspire future technological advancements, it also assesses the impact of technological change on

[24] Clarke, "Asimov's Laws of Robotics (Part II)," 61.
[25] Clarke, "Asimov's Laws of Robotics (Part II)," 61.
[26] Clarke, "Asimov's Laws of Robotics (Part II)," 62 – 64.

humans. Works of science fiction compel the reader to grapple with the effects of science on society.[27] These works assess the ramification of technological advances and become more important for their thought-provoking questions rather than their visions of technological wizardry. This focus on dilemmas and questions rather than technology helps explain the continuing relevance of works once society and contemporary technology have marched past the technology proffered in the work.[28] For example, Mary Shelley's 1823 novel *Frankenstein* raises fears of human creations rising up and turning against their creator.[29] Additionally, H.G. Wells's 1898 work *War of the Worlds* has maintained continuing relevance as directors adapt the perceived existential threat of the time to Wells's vision.[30] In addition to the 1938 radio broadcast by Orson Wells, adaptations of the work have echoed fears of nuclear war during the Cold War and the threat of terrorism after the 11 September 2001 attacks.[31]

This type of critique exists in Asimov's works. Elements of contemporary arguments against advanced forms of artificial intelligence emerge in *I, Robot*. Arguing with her husband, George, about whether they should allow Robbie the robot to watch their daughter, Mrs. Weston emphatically states she will not have her daughter "entrusted to a machine" because it "has no soul, and no one knows what it may be thinking."[32] A "Society for Humanity" arises and argues the machine "robs man of soul."[33] Army Major Daniel Davis's argument against autonomous systems in his article "Who Decides: Man or Machine?"

---

[27] P.W. Singer, *Wired for War: The Robotics Revolution and Conflict in the 21st Century* (New York: Penguin, 2009), 153.
[28] Singer, *Wired for War*, 153.
[29] Arkin, *Governing Lethal Behavior in Autonomous Robots*, 56. Mary Shelley, *Frankenstein* (New York: SoHo Books, 2010). The first edition of *Frankenstein* appeared anonymously in 1818. Shelley's name appears on the second addition published in 1823.
[30] H.G. Wells, *War of the Worlds* (New York: SoHo Books, 2010).
[31] Singer, *Wired for War*, 153.
[32] Asimov, *I, Robot*, 25.
[33] Asimov, *I, Robot*, 213.

echoes strains of the arguments by Mrs. Weston and the Society for Humanity. Davis describes the potential of creating "an efficient, heartless killing machine to be feared and hated."[34] Thus, the dilemmas Asimov foresaw remain.

As a work of science fiction, Asimov's Three Laws of Robotics provide a starting point for the discussion on rules to govern robot, or autonomous system, behavior. The Three Laws, however, do not provide the answer for governing such behavior. Inherent dilemmas exist in Asimov's formulation. While the Three Laws serve a "useful fictional purpose," they are "at best a straw man to bootstrap the ethical debate."[35] Expanding on Asimov's vision, engineers have begun developing schema for implementing a machine ethic.

Basic Approaches for Developing a Robot Ethic

Developing a universal robot ethic is a daunting task. The challenge facing engineers developing an ethical system for military systems is (arguably) less complex because the law of armed conflict and rules of engagement bound the set of behaviors.[36] The two basic approaches for developing the ethical system to guide an autonomous system are a top-down approach and a bottom up approach.[37] A top-down approach would encode a particular ethical theory. The system would utilize the ethical theory to rank possible options for moral acceptability. Conversely, systems utilizing a bottom-up approach would develop or learn morality based on experience, much as children do as they mature into adults.[38]

---

[34] Maj Daniel L. Davis, "Who Decides: Man or Machine?" *Armed Forces Journal*, November 2007, http://www.armedforcesjournal.com/2007/11/3036753 (accessed 16 February 2012).
[35] Arkin, *Governing Lethal Behavior in Autonomous Robots*, 48.
[36] Lin, *Autonomous Military Robotics*, 25.
[37] Lin, *Autonomous Military Robotics*, 27.
[38] Lin, *Autonomous Military Robotics*, 27.

A deontological approach to ethics, which presents a set of inflexible rules, is one means to implement a top-down approach.[39] Behavior in accordance with the rules is moral, and breaking the rules is immoral.[40] The system would know the rules, or how to determine the rules, and could apply the rules to specific circumstances.[41] Immanuel Kant's Categorical Imperative typifies a deontological approach to ethics.[42] Asimov's Three Laws of Robotics is another deontological approach. An inherent limitation with a deontological approach emerges when a conflict between the rules exists. Context and exceptions to the rules matter. If the rules are not comprehensive enough, there is no guarantee of proper behavior.[43]

Slave morality is a problem when implementing deontological ethics. A system slavishly following commands would not exhibit true Kantian autonomy. While the system could make autonomous decisions about the means to execute a pre-programmed goal, it could not make autonomous decisions about the goals themselves. The question of ethics would then collapse to the ethics of the person assigning the goals–the commander in a military context. The problem with slave morality is that

---

[39] As a moral theory, deontology guides and assesses an actor's choices by what the actor ought to do as opposed to who or what the actor should be. According to deontology, what makes a choice correct is its adherence to a norm. An actor, therefore, cannot justify some choices regardless of the consequence's moral goodness. The three major approaches in normative ethics are deontology, consequentialism, and virtue ethics. For more information, see Larry Alexander and Michael Moore, "Deontological Ethics," *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*, ed. Edward N. Zalta, http://plato.stanford.edu/archives/fall2008/entries/ethics-deontological (accessed 20 February 2012).

[40] Lin, *Autonomous Military Robotics*, 29.

[41] Wendell Wallach and Colin Allen, *Moral Machines: Teaching Robots Right from Wrong* (New York: Oxford University Press, 2009) 85 – 86.

[42] The Categorical Imperative has two primary components. The first component, known as the formula of universal law, commands, "Act only in accordance with that maxim through which you can at the same time will that it become a universal law." The second component is the formula of the end in itself, which directs, "Act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means. See Immanuel Kant, *Grounding for the Metaphysics of Morals*, ed. and trans. James W. Ellington (Indianapolis, IN: Hackett Publishing, 1993), 4:421 and 4:429.

[43] Lin, *Autonomous Military Robotics*, 31 – 32.

"consequences matter morally, and simply following the rules is morally wrong if it leads to bad outcomes."[44]

Utilitarianism, a consequentialist approach to ethics, provides another top-down approach.[45] Rather than focusing on the rigid adherence to rules and duties, utilitarianism emphasizes the importance of outcomes resulting from action. Utilitarianism seeks the greatest good for the greatest number. For an autonomous system, however, utilitarianism faces calculation problems. Issues include how to represent utility for a system and how broadly to analyze the consequences of action. Due to the significant computational challenges associated with a utilitarian calculation, even the fast systems might not be able to determine the most acceptable course of action in a timely manner. An inability to calculate utility would obviate the value of utilitarianism in an autonomous system.[46]

While top-down approaches help guide humans in their evaluations of potential courses of action, implementing a top-down approach in an autonomous system appears problematic. The rules of a top-down approach do not always provide unequivocal guidance. A top-down approach often requires goals that are defined either "so vaguely and abstractly that their meaning and their application to specific situations is debatable, or they are defined so rigidly that they fail to produce decisions that are appropriately sensitive to new context."[47] Furthermore, reaching a consensus among humans about which specific rules are consistent with "über-rules," such as Kant's Categorical

---

[44] Lin, *Autonomous Military Robotics*, 33.

[45] Consequentialism is the perspective that consequences determine normative properties. With a consequentialist viewpoint, the moral rightness of an act depends only the consequences of the act, or something related to the act, such as the intent behind the act. For more information see, Walter Sinnott-Armstrong, "Consequentialism," *The Stanford Encyclopedia of Philosophy (Winter 2011 Edition)*, ed. Edward N. Zalta, http://plato.stanford.edu/archives/win2011/entries/consequentialism (accessed 20 February 2012).

[46] Lin, *Autonomous Military Robotics*, 34.

[47] Lin, *Autonomous Military Robotics*, 34.

Imperative, is difficult.[48] Moral reasoning is a complex task. It requires the ability to distinguish the letter of the law from the spirit of the law, which emerges from experience.[49] This ability to learn from experience gives rise to bottom-up approaches.

A system with a bottom-up approach to ethical reasoning learns appropriate behavior through experience. The processes of performance optimization, evolution, and human learning and development have inspired bottom-up approaches.[50]

A performance optimization technique utilizes trial-and-error to progressively fine tune behavior. This process eventually enables the system to meet or exceed task performance criteria. With performance optimization techniques, the resultant behavior can achieve a high level of performance even if the design engineers do not know the best way to decompose the task into subtasks. *Post hoc* analysis may lead to an understanding of how the subtasks produce the results, but this analysis typically does not correlate to the task decomposition developed during *a priori* analysis.[51]

Evolution provides another model for system optimization based on performance criteria that utilize self-selection and self-organization. The successful agents within the system perform certain tasks of interest better than the other agents. The successful agents undergo a modification and recombination process analogous to sexual reproduction to produce a new generation of agents. The process then repeats with the new generation.[52] As the system evolves, each new generation of agents is more capable of performing the task of interest. Thus, system learning emerges through the evolution process.

---

[48] Wallach and Allen, *Moral Machines*, 96 – 97.
[49] Wallach and Allen, *Moral Machines*, 97.
[50] Lin, *Autonomous Military Robotics*, 34 – 35.
[51] Lin, *Autonomous Military Robotics*, 35.
[52] Lin, *Autonomous Military Robotics*, 35 – 36.

Alan Turing first suggested using the moral development of a child for the creation of machine intelligence.[53] In his influential article "Computing Machinery and Intelligence," Turing explores the question "Can machines think?" To assess this question, he proposes an imitation game, the eponymous Turing test, to assess whether a machine can think.[54] To develop this machine, Turing suggests, "Instead of trying to produce a programme to simulate the adult mind…try to produce one which simulates the child's."[55] Indeed, human wisdom develops "from experience, from attentive doing and observing, from the integration of cognition, emotions, and reflection."[56] While the moral development of a child provides a useful model for machine learning, current algorithms and techniques are not robust enough for implementation.[57]

The strength of bottom-up approaches is the ability to produce dynamic morality, which facilitates varied responses as conditions change. Bottom-up approaches are a form of ethical reasoning known as particularism.[58] The particularism approach recognizes each ethical situation as unique and asserts that general laws are not possible.[59] As context and circumstances change, however, determining which goals to utilize for evaluating potential courses of actions may become difficult.

---

[53] Lin, *Autonomous Military Robotics*, 36.

[54] In Turing's imitation game, a human judge attempts to determine whether a conversation is between two humans or between a human and a machine. A machine passes the test if the judge thinks the conversation between the machine and the human is between two humans. See Alan M. Turing, "Computing Machinery and Intelligence," *Mind: A Quarterly Review of Psychology and Philosophy* 49, no. 236 (October 1950), 433 – 434.

[55] Turing, "Computing Machinery and Intelligence," 456.

[56] Wallach and Allen, *Moral Machines*, 97.

[57] Lin, *Autonomous Military Robotics*, 36.

[58] The extreme version of moral particularism holds that defensible moral principles do not exist. Variation is a key feature of particularism. While a generalist demands sameness in how considerations function across cases, a particularist allows other aspects of the case to influence whether the feature is relevant to the case. For more information, see Jonathan Dancy, "Moral Particularism," *The Stanford Encyclopedia of Philosophy (Spring 2009 Edition)*, ed. Edward N. Zalta, http://plato.stanford.edu/archives/spr2009/entries/moral-particularism (accessed 20 February 2012).

[59] Lin, *Autonomous Military Robotics*, 36.

Bottom-up approaches are most effective when directed toward one clear goal. In a situation with multiple or unclear goals, bottom-up approaches are less likely to provide a clear course of action.[60]

A particular concern with learning systems is the possibility that the system could devise a method to override any control mechanisms restraining system behavior.[61] Machines could run amok. In *The Terminator*, this scenario enables Skynet to become self-aware and turn against humanity.[62]

A hybrid approach to constructing a moral reasoning capability for a machine would integrate top-down and bottom-up approaches. It would also incorporate supra-rational mechanisms, which enable understanding of social context by interpreting emotions and other forms of non-verbal communication. Humans utilize a hybrid approach to moral reasoning. Evolution and learning shape bottom-up mechanisms for moral judgment, while a top-down approach enables reasoning about ethical challenges.[63]

Virtue ethics provides a potentially viable hybrid approach for developing moral judgment in a machine.[64] It focuses on character, not rules. Under virtue ethics, actions reveal morality rather than constituting an agent's morality. The proper question to ask about morality shifts away from which rule to follow or which rules apply to this act to "what sort of person will I reveal myself to be (or become) if

---

[60] Lin, *Autonomous Military Robotics*, 37.
[61] Wallach and Allen, *Moral Machines*, 195.
[62] James Cameron, *The Terminator* (Los Angeles, CA: MGM, 1984).
[63] Lin, *Autonomous Military Robotics*, 37 – 38.
[64] Virtue ethics is a normative approach to ethics emphasizing virtue, or moral character. As a normative approach to ethics, virtue ethics stands in contrast to deontology, which emphasizes rules, and consequentialism, which emphasizes consequences. The three central concepts of virtue ethics are virtue, practical wisdom, and *eudaimonia*, which translated from Greek means happiness, flourishing, or well-being. For more information, see Rosalind Hursthouse, "Virtue Ethics," *The Stanford Encyclopedia of Philosophy (Winter 2010 Edition)*, ed. Edward N. Zalta, http://plato.stanford.edu/archives/win2010/entries/ethics-virtue (accessed 20 February 2012).

this is the sort of thing I do."[65] Since virtue ethics are context dependent, a single rule or set of rules cannot specify the actions various agents in different rules ought to perform.[66]

In a sense, virtue ethics combines both bottom-up and top-down approaches. The development of virtues results from bottom-up learning, while the virtues themselves provide a top-down mechanism to evaluate the actions of the system. As virtues develop through learning and experience with reinforcement of positive results, accumulated data enables the system to produce generalized responses that exceed specific training. These emergent learned patterns, however, do not have an accompanying explanation for why the system chose the action. A mechanism for evaluating and providing explanations for the learned behavior could arise by combing the bottom-up development behavior with a top-down implementation of virtues.[67]

Hybrid architectures appear to represent the preferred method for developing the moral reasoning capability for autonomous systems. This methodology mimics the means by which humans gain ethical experience.[68] In the course of executing their mission, autonomous systems will need to make decisions. Since lethal autonomous systems will make decisions with harmful consequences, an ethically blind system is not desirable. The belief that lethal autonomous systems "will honor basic human values and norms in their choice of actions" will provide the foundation for societal acceptance of these systems.[69] Furthermore, military commanders will need to have the confidence that the lethal autonomous systems will only engage appropriate targets.[70]

While any of these approaches may eventually provide a moral reasoning ability for autonomous systems capable of employing lethal

---

[65] Lin, *Autonomous Military Robotics*, 39.
[66] Lin, *Autonomous Military Robotics*, 39.
[67] Lin, *Autonomous Military Robotics*, 40.
[68] Lin, *Autonomous Military Robotics*, 41.
[69] Lin, *Autonomous Military Robotics*, 41.
[70] Lin, *Autonomous Military Robotics*, 41.

force, thinking about the challenges facing designers of such systems will help ensure the placement of adequate system safeguards.[71]

Engineering Techniques for Implementation

To implement these basic approaches, engineers must capture relevant aspects of the human cognitive process. Joanne Thoms of Bath University and BAE Systems provides an engineering framework focusing on the three cognitive capabilities of awareness, understanding, and deliberation and the associated information flow.[72] Figure 5 illustrates Thoms's framework.

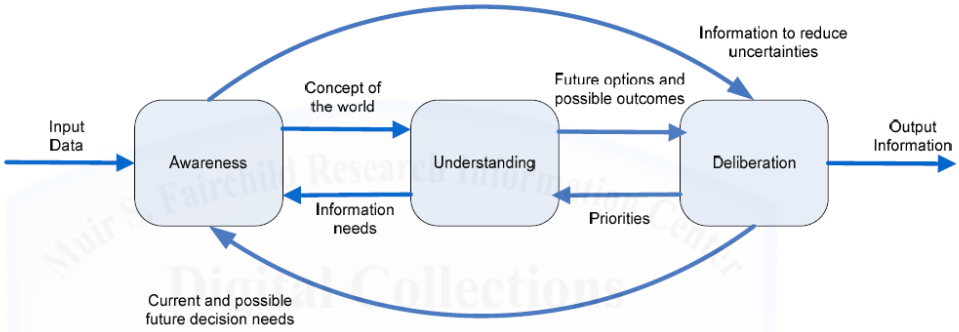

Figure 5: Thoms's Cognitive Framework

Source: Joanne Thoms, "Understanding the Impact of Machine Technologies on Human Team Cognition," 3

John Boyd's Observe, Orient, Decide, Act (OODA) Loop, shown in figure 6, also captures the human cognitive process.

[71] Wallach and Allen, *Moral Machines*, 172.
[72] Joanne Thoms, "Understanding the Impact of Machine Technologies on Human Team Cognition," *4th Systems Engineering for Autonomous Systems Defense Technology Centre (SEAS DTC) Technical Conference* (Edinburgh, UK: Defense Technology Centres, 7 – 8 July 2009): Paper B7, 2.

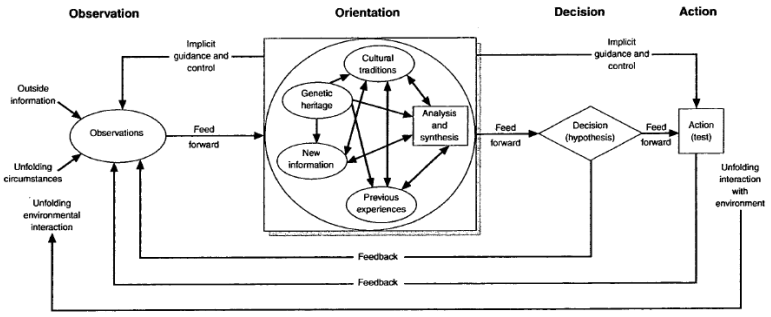

Figure 6: Boyd's OODA Loop

Source: Frans Osinga, Science, Strategy and War, 231

Another common cognitive model, typically utilized in the fields of automation and robotics, is the sense, perceive, decide, and act model. These four steps address the acquisition of information, the analysis of information, the selection of a decision based on this information, and the execution of the decision.[73] Thus, the sense, perceive, decide, act model is roughly analogous to the four steps of Boyd's OODA loop.

Cognitive models provide a methodology to convert the human role into the decision process of an autonomous system.[74] These models also provide a mechanism to break down complex tasks associated with cognition into discrete elements. Engineers can utilize these cognitive models to develop requirements for the cognitive capabilities as functions, inputs, and outputs. Breaking down the cognitive capabilities in this manner aids engineers in assessing whether a particular technology will meet system requirements.[75]

While engineers strive to deconstruct the human cognitive process, uncertainty still confronts the decision-making process of an

---

[73] Raja Parasuraman, Thomas B. Sheridan, and Christopher D. Wickens, "A Model for Types and Levels of Human Interaction with Automation," *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans* 30, no. 3 (May 2000), 286 – 287.
[74] Tony Gillespie and Robin West, "Requirements for Autonomous Unmanned Air Systems Set by Legal Issues," *The International C2 Journal* 4, no. 2 (2010): 15 – 16.
[75] Gillespie and West, "Requirements for Autonomous Unmanned Air Systems Set by Legal Issues," 17.

autonomous system. As the Prussian General Helmuth von Moltke asserted, "No plan of operations extends with certainty beyond the first encounter with the enemy's main strength."[76] Markov decisions processes, named for the Russian mathematician A.A. Markov, provide a method to formalize planning for problems under conditions of uncertainty.[77] Markov developed the concept of chain dependence in a random, or stochastic, process.[78] He was also the first to systematically study a particular type of stochastic process in which a system's state is independent of its past. In these processes, now eponymously known as Markov processes, the system's future state depends only on the present state, not the preceding sequence of events.[79] Markov decision processes extend Markov processes into decision making under uncertainty.

In a decision-making process, uncertainty leads to the decision-maker and random events each having partial control over the outcome. Markov decision processes provide a mathematical framework to model decision-making problems with uncertainty, utilizing both probability theory and utility theory. The process employs a dynamic world model and reward structure to ascertain an optimal set of actions for implementation to achieve a goal.[80] With uncertainty, small changes in the data create the potential to make possible solutions unfeasible. A

---

[76] Helmuth Graf von Moltke, *Moltke on the Art of War: Selected Writings*, trans. Daniel Hughes (New York: Ballantine, 1993), 45.

[77] Nalan Gulpinar and Ethem Canakoglu, "Robust Team Coordination and Decision Making under Uncertainty," *5th SEAS DTC Technical Conference* (Edinburgh, UK: Defense Technology Centres, 14 – 15 July 2010): Paper B6, 1.

[78] A stochastic process, as defined by the American mathematician J.L. Doob, is the mathematical abstraction of an empirical process whose development is governed by probabilistic laws. See, A.T. Bharucha-Reid, *Elements of the Theory of Markov Processes and Their Applications* (Mineola, NY: Dover Publications, 1960), 3 – 4. A stochastic process is more commonly known as a random process. Given a set of known initial conditions, a system following a stochastic process has the potential to evolve into many different outcomes. In contrast, a deterministic process does not involve randomness. Thus, given a set of known initial conditions, a system following a deterministic process will always produce the same outcome.

[79] Bharucha-Reid, *Elements of the Theory of Markov Processes and Their Applications*, 3 – 4.

[80] Gulpinar and Canakoglu, "Robust Team Coordination and Decision Making under Uncertainty," 1.

robust solution, therefore, inoculates the model's performance against uncertainty effects by determining the optimal decision with the worst possible values of uncertain parameters. Performance improves if these worst-case parameters do not occur.[81]

With a team of systems, such as a swarm of autonomous systems performing a reconnaissance mission, coordination is necessary for effective and efficient resource utilization. Under these circumstances, the autonomous decision-making process allocates tasks and resources to maximize overall team performance.[82] While coordination among the systems is necessary for resource allocation, uncertain task duration emerges as part of the coordination problem, yet the system must complete all tasks within the prescribed time limits. Markov decision-making processes provide a method to cope with these uncertainties, enabling the coordination of a team of autonomous decision-making systems.[83]

Within a decision-making process, standard operating procedures and rules of behavior, such as the law of armed conflict, constrain possible behavior. Engineers represent these behavior constraints as norms. Soft constraints are those constraints an agent may potentially violate. For example, under the general rules of engagement, the soft constraint emerging from the prohibition against firing at another nation's soldiers dissolves under conditions of self-defense after demonstrated hostile acts and intent. These soft constraints subsequently oblige, permit, or prohibit certain actions by an agent under particular conditions.[84] In the preceding example, the

---

[81] Gulpinar and Canakoglu, "Robust Team Coordination and Decision Making under Uncertainty," 5.
[82] Gulpinar and Canakoglu, "Robust Team Coordination and Decision Making under Uncertainty," 6.
[83] Gulpinar and Canakoglu, ""Robust Team Coordination and Decision Making under Uncertainty,"11.
[84] Nir Oren, Simon Miles, and Michael Luck, "Representing Norms within Agent Systems," *5th SEAS DTC Technical Conference,* (Edinburgh, UK: Defense Technology Centres, 14 – 15 July 2010): Paper B2, 1.

demonstration of hostile act and intent permits, but does not require, the use of force.

Conditional norms do not always affect the agent and require particular conditions for activation. Accordingly, the agent must have the capability to assess the environment in order to determine which norms apply. Through this process, the system accomplishes any obligations, complies with prohibitions, and evaluates permissible actions.[85] Activation conditions identify when a norm applies, and expiration conditions determine when a norm no longer applies.[86] For instance, an activation condition would trigger the norm for the use of countermeasures when passing within a certain distance of a threat while the expiration condition would specify when to stop utilizing the countermeasures. A complete norm would consist of the following components: type identifier (obligation, prohibition, or permission), activation condition, norm condition to identify the desired end state, an expiration condition, and a norm target, which are the agents that the norm affects.[87]

An associate system, or expert system, is a method of combing norms for a decision-making process. Associate systems mimic human decision-making by utilizing an operationally tailored knowledge base to reason about mission objectives, deduce the intentions of others, and develop potential courses of action.[88] The reasoning capability of the associate system enables it to make actionable recommendations, to direct other components and subsystems, and to execute the mission automatically, if necessary.[89]

---

[85] Oren, Miles, and Luck, "Representing Norms within Agent Systems," 2 – 3.
[86] Oren, Miles, and Luck, "Representing Norms within Agent Systems," 4.
[87] Oren, Miles, and Luck, "Representing Norms within Agent Systems," 4.
[88] Sheila B. Banks and Carl S. Lizza, "Pilot's Associate: A Cooperative, Knowledge-Based System Application," *IEEE Intelligent Systems and their Applications*, June 1991, 18 – 29.
[89] Sarah Johnston, Roy Sterritt, Edward Hanna, and Patricia O'Hagan, "Reflex Autonomicity in an Agent-Based Security System: The Autonomic Access Control

Several technological limitations with the potential to hinder the development of associate technology exist. The unforeseen and often undesirable consequences of programming bugs in software necessitate vigorous testing during the development process to find and correct the bugs. Complicating the programming task, human decision-making processes are messy and not fully understood, making translation to programming rules difficult. Yet, an associate system must make decisions in a rational and consistent way.[90] Another complicating factor stems from the complexity of modeling nonlinear behavior.[91] The susceptibility of models to slight variations in initial conditions also creates difficulties.[92] While these issues will present challenges to the development of an associate type system, they are not barriers to development but rather only impediments.[93]

Engineers have begun developing software architectures to implement these basic approaches and techniques, which is the next area of investigation.

System," *Fourth IEEE International Workshop on Engineering of Autonomic and Autonomous Systems*, March 2007, 1.

[90] William B. McClure, *Technology and Command: Implications for Military Operations in the Twenty-first Century*, Occasional Paper No. 15 (Maxwell AFB, AL: Center for Strategy and Technology, Air War College, Air University, July 2000), 25

[91] McClure, *Technology and Command: Implications for Military Operations in the Twenty-first Century*, 25. A linear system adheres to the superposition principle, by which the net response of a system caused by multiple stimuli is the sum of the individual responses caused by each stimuli. In a nonlinear system, the solution is not a combination of the individual components. Nonlinear behavior can exhibit indeterminism (cannot predict system behavior), multi-point stability (system behavior alternates between two or more states), or aperiodic oscillations (chaotic behavior).

[92] On 25 February 1991, a Patriot missile battery operating at Dhahran, Saudi Arabia during Operation DESERT STORM failed to track and intercept an incoming Scud missile due to a software timing error. At the time of the incident, the system's computer had been operating continuously for 100 hours. The software timing inaccuracy was 0.3433 seconds, which caused range gate shift of the system's radar of approximately 687 meters, which was significant enough to prevent the system from detecting the incoming missile. For more information, see *Patriot Missile Software Problem*, GAO/IMTEC-92-26 (Washington, DC: General Accounting Office, 1992).

[93] McClure, *Technology and Command: Implications for Military Operations in the Twenty-first Century*, 25.

Arkin's Ethical Architecture

Georgia Institute of Technology Robotics Professor Ronald Arkin has proposed the architecture for implementing ethical decision-making in a machine. Arkin argues engineers could encode the laws of armed conflict to effectively program compassion and mercy into an autonomous system. Consequently, he believes the resultant force would be a more humane force than a human force.[94]

Arkin utilizes formalisms to discern relationships necessary to support moral reasoning within the architecture design of autonomous systems. Mathematically describing the relationship between sensing and action with functional notation produces the expression:

$$\beta(\boldsymbol{s}) \xrightarrow{yields} \boldsymbol{r}$$

The behavior β, given stimulus **s**, produces response **r**.[95] From this general formula, Arkin utilizes formalisms to describe the relationships within the system architecture necessary to support ethical reasoning.

Of the set of entire behavior responses, lethal behaviors comprise a subset. Ethical lethal behaviors are a further subset of lethal behaviors. Figure 7 illustrates the nesting of permissible behaviors into subsets given a particular situational context.

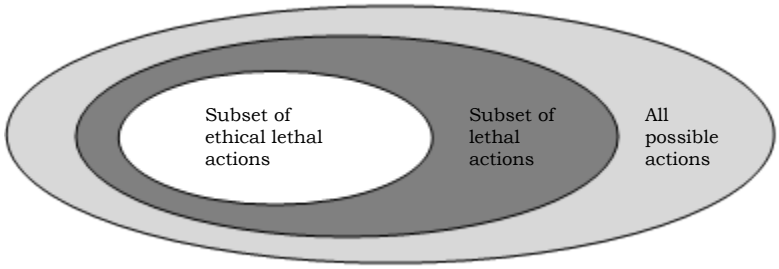

Figure 7: Behavioral Action Space

Source: Adapted from Ronald Arkin, Governing Lethal Behavior in Autonomous Robots, 65

[94] Arkin, *Governing Lethal Behavior in Autonomous Robots*, 46.
[95] Arkin, *Governing Lethal Behavior in Autonomous Robots*, 57.

The system coordinating function provides a given set of ethical constraints. According to the closed world assumption, the system presumes anything not currently known as true is therefore false. If the system encounters a situation outside coordinating function, then based on the closed world assumption it cannot generate a lethal response. This characteristic prevents the system from applying lethal force in situations not governed, or outside of, given ethical constraints.[96]

From this, Arkin develops five functions for the controller design.

1. The *ethical situation requirement* ensures only certain situations governed by the controller can result in a lethal action. Situations outside this behavior set cannot result in the application of lethal force.
2. The *ethical response requirement* ensures, given a specific situation, only permissible actions result.
3. The *unethical response prohibition* inhibits unethical responses from occurring, modified unethical responses into ethical responses through modifications such as removing the application of force, or precludes the generation of unethical responses through architecture design.
4. The *obligated lethality requirement* confirms the existence of at least one constraint from the rules of engagement obligating the use of lethality in the given circumstance before producing a lethal response.
5. The *jus in bello compliance* verifies adherence to proportionality requirements and combatant/non-combatant discrimination requirements.[97]

From ethical reasoning, a particular action may be obligatory, permissible, or forbidden. As a safeguard, Arkin argues lethal actions executed by an autonomous system must be obligatory, not solely

[96] Arkin, *Governing Lethal Behavior in Autonomous Robots*, 63.
[97] Arkin, *Governing Lethal Behavior in Autonomous Robots*, 65 – 66.

permissible. He uses the laws of armed conflict and rules of engagement to determine forbidden lethal actions and turns to the rules of engagement to derive mission requirements necessitating the use of lethal force.[98]

Based upon requirements derived from these conditions, Arkin proposes a specific ethical architecture with four components–an ethical governor, an ethical behavior control, an ethical adaptor, and a responsibility advisor. The ethical governor would provide the constraints on permissible behavior and force the system to form a second opinion prior to employing lethal force. The ethical behavior control would ensure that lethal responses only occurred within acceptable ethical constraints. The ethical adaptor would review the system's actions after the employment of lethal force. If the ethical adaptor determined the behavior of the system violated the Laws of War or Rules of Engagement, it would further constrain the system's behavior to prevent repetition of the violation. Finally, the responsibility advisor would be the interface between the human operator and the lethal autonomous system. The responsibility advisor would ensure the human operator understood the scenarios under which the system would employ lethal force and have the human operator explicitly authorize the use of the lethal autonomous system.[99] With Arkin's architecture, before the autonomous system may act with lethal force, an operator must accept responsibility. The system must then establish a case for military necessity, maximize target discrimination, and finally minimize the force applied by examining proportionality and by applying the principle of double intention.[100] According to the principle of double intention, soldiers must assume mission risk to avoid killing civilians. The limit of these risks arises at the

---

[98] Ronald Arkin, *Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture,* Technical Report GIT-GVU-07-11 (Atlanta, GA: Georgia Institute of Technology, 2008), 40.
[99] Arkin, *Governing Lethal Behavior in Autonomous Robots*, 125 – 126.
[100] Arkin, *Governing Lethal Behavior in Autonomous Robots*, 121.

point where additional risk-taking would doom the mission or make it so costly that the military could not re-attempt the mission.[101] Figure 8 provides a graphical representation of Arkin's architecture.

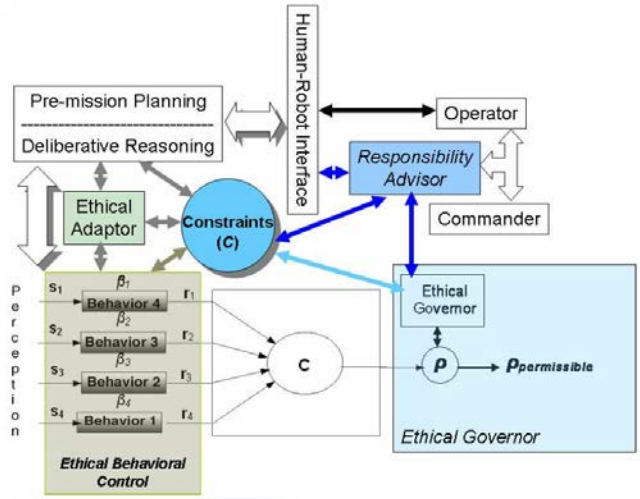

Figure 8: Arkin's Ethical Decision-Making Architecture

Source: Ronald Arkin, Governing Lethal Behavior, 62

The ethical governor would accomplish evidential reasoning and constraint application processes. The evidential reasoning processes would interpret incoming situational awareness data as evidence to reason about lethal behavior. The constraint application process would utilize the evidence to apply ethical constraints derived from the law of armed conflict and rules of engagement. These ethical constraints would suppress possible unethical behavior.[102] Figure 9 illustrates the ethical governor.

[101] Michael Walzer, *Just and Unjust Wars: A Moral Argument with Historical Illustrations* (New York: Basic Books, 1977), 157.
[102] Arkin, *Governing Lethal Behavior in Autonomous Robots*, 178 – 179.

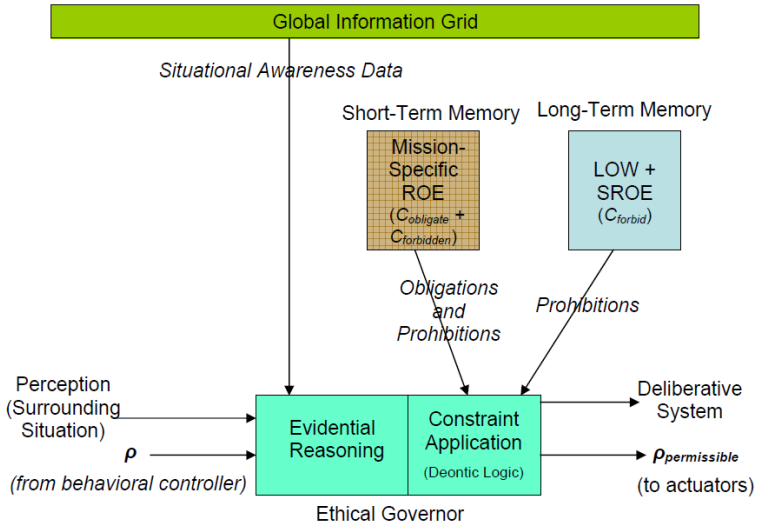

Figure 9: Arkin's Ethical Governor

Source: Ronald Arkin, Governing Lethal Behavior, 66

A proportionality algorithm would conduct an assessment as part of the ethical governor. The proportionality algorithm would evaluate weapons and weapons release positions for ethical constraints. It would also evaluate these factors for collateral damage given the military necessity of the specific target. If the proportionality algorithm identified a weapon and weapon release position combination satisfying ethical constraints while minimizing collateral damage in relation to the military necessity of the target, the proportionality algorithm would advise that lethal force would be permissible in this situation.[103]

The ethical behavior control would provide a mechanism to produce ethical behavior consistent with the laws of armed conflict and rules of engagement rather than relying on the ethical governor to restrict behavior. Ideally, the set of lethal actions the system could produce would lie in the ethical lethal behavior subset.[104]

The ethical adaptor would provide a mechanism to compensate for potential system errors regarding the use of lethal force. This mechanism

[103] Arkin, *Governing Lethal Behavior in Autonomous Robots*, 187.
[104] Arkin, *Governing Lethal Behavior,* 69 – 70.

would enable both after-action and real-time restrictions of lethal behavior. The after-action reflection would evaluate the system's mission performance. The system's post-mission internal affective state, represented by guilt or remorse, could trigger the after-action reflection. Based on the guidance resulting from this analysis, the system architecture would modify its parameters, effectively altering the system's ethical basis to promote proper action in the future. Conversely, the real-time restriction of lethal behavior would occur if the system exceeded threshold values of the affective state, again represented by guilt or remorse. In this case, the system would immediately cease employing lethal force.[105]

The responsibility advisor would enable a human to provide mission authorization. During this process, the operator would acknowledge proper training on the use of the system and acknowledge system obligations authorizing the use of force.[106] The Responsibility Advisor would also enable a system override capability. If the operator were directly controlling the system, it would still advise the operator on ethical constraint violations. In this event, the system could implement a two-trigger pull before applying lethal force. The system would first warn the operator of perceived ethical violations. The operator would have to confirm responsibility for the override before the system would employ lethal force.[107] The opposite override inhibiting the employment of lethal force would not require a two-trigger pull process. The operator would have the capability to override and stop pending actions as a form of an emergency stop for the system.[108]

Arkin specifically designed this architecture to adhere to the law of armed conflict principles.[109] By doing so, he allays some of the moral and

---

[105] Arkin, *Governing Lethal Behavior in Autonomous Robots*, 138.
[106] Arkin, *Governing Lethal Behavior in Autonomous Robots*, 148 – 149.
[107] Arkin, *Governing Lethal Behavior in Autonomous Robots*, 151.
[108] Arkin, *Governing Lethal Behavior in Autonomous Robots*, 152.
[109] Arkin, *Governing Lethal Behavior in Autonomous Robots*, 97.

ethical criticisms of lethal autonomous systems. Inclusion of the responsibility advisor in the design addresses accountability concerns, but it also raises concerns related to applying force at a distance and automation bias explored in the next chapter. Furthermore, his proposed architecture still depends on technical developments to address concerns about discrimination and accountability.

Machines Targeting Machines

John Canning of the US Navy's Naval Surface Warfare Division offers an alternative method for achieving an autonomous weapons employment capability. Noting the requirement for new weapons to undergo a legal review prior to full-scale production, Canning reasoned legal considerations would be the "single biggest issue"–a reverse salient of significant proportion.[110] The emerging legal concern was the possibility these weapons might not limit civilian casualties and damage. The military lawyers with whom Canning spoke did not have the same concerns with systems "that could automatically target 'things,' instead of 'people.'"[111] Canning noted the existence of weapons that did not directly target people and a metaphorical acceptance of targeting the archer's bow or arrows rather than the archer. For example, the Aegis combat system on Ticonderoga class guided missile cruisers has a special automatic mode to protect the ship in anti-air warfare. Additionally, an old anti-ship variant of the Tomahawk cruise missile had the capability to automatically identify and attack specific enemy ships

[110] John S. Canning, "Weaponized Unmanned Systems: A Transformational Warfighting Opportunity, Government Roles in Making it Happen," *Engineering the Total Ship Symposium* (Falls Church, VA: American Society of Naval Engineers, 23 – 25 September 2008), 3. The notion of a reverse salient in a technical system describes a system component that has fallen behind or is out of phase with other system components. For additional information on the concept of a reverse salient in a technical system, see Thomas P. Hughes, "The Evolution of Large Technological Systems" in *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*, eds. Wiebe E. Bijker, Thomas P. Hughes, and Trevor Pinch (Cambridge, MA: MIT Press, 1989), 71 – 76.

[111] Canning, "Weaponized Unmanned Systems," 4.

after launch.[112] Furthermore, Canning argues, international treaties ban anti-personnel mines but not anti-tank mines because the former target people while the latter target machines. Targeting machines instead of people, however, does not obviate concerns with collateral damage.[113]

This approach restricts the targets of an autonomous system to a subset of valid targets. Under certain circumstances, both people and things, such as military equipment, buildings, and weapons, are valid, legal targets. When people mix with things, targeting requires discretion and judgment. Table 3 presents targeting considerations with a human exercising judgment on whether to apply lethal force.

Table 3: Targeting Considerations for Humans Applying Lethal Force[114]

		Target Set: People	
		Not a Valid Military Target	Valid Military Target
Target Set: Things (e.g. equipment, buildings, weapons, etc.)	Not a Valid Military Target	Cannot Target	Target People, Not Things with Collateral Damage Estimates
	Valid Military Target	Target Things, Not People With Proportionality Considerations	Target People and Things

Source: Adapted from John S. Canning, "A Concept of Operations for Armed Autonomous Systems," 15.

Table 4 presents the targeting subset with an autonomous system exercising judgment on the application of lethal force.

---

[112] Canning, "Weaponized Unmanned Systems," 4.
[113] Canning, "Weaponized Unmanned Systems," 7.
[114] Adapted from John S. Canning, "A Concept of Operations for Armed Autonomous Systems," *3rd Annual Disruptive Technology Conference "Seeking the Capability Before the Capability is the Surprise,"* (Washington, DC: National Defense Industrial Association, 6 – 7 September 2006), 15.

Table 4: Targeting Considerations for Autonomous Systems Applying Lethal Force[115]

		Target Set: People	
Target Set: Things (e.g. equipment, buildings, weapons, etc.)		Not a Valid Military Target	Valid Military Target
	Not a Valid Military Target	Cannot Target	Prohibited from Targeting by Design
	Valid Military Target	Target Things, Not People with Proportionality Considerations	Target Things, Not People by Design

Source: Adapted from John S. Canning, "A Concept of Operations for Armed Autonomous Systems"

To implement Canning's approach, automatic target recognition capabilities and sensors become key enabling technologies.[116] The reliance of autonomous systems on sensors for targeting information creates a vulnerability to deception. The autonomous system, therefore, must have an alternative targeting method to overcome the adversary's deception strategies.[117]

Canning's approach of autonomous systems targeting things rather than people presents an interesting legal sleight-of-hand. This approach blunts some criticism of autonomous systems by the emphasis of its targeting technique, because this methodology of targeting the archer's bow rather than the archer has historical precedent.[118] The automatic mode of the Aegis combat system on Ticonderoga class cruisers, the CATPOR mine which automatically fires a torpedo at detected submarines, and the general acceptance of anti-tank mines but

---

[115] Adapted from Canning, "A Concept of Operations for Armed Autonomous Systems," 16.
[116] Canning, "Weaponized Unmanned Systems," 5. Also see, Paul S. Canning, *A Definitive Work on Factors Impacting the Arming of Unmanned Vehicles*, NSWCDD TR-05/36 (Dahlgren, VA: Department of the Navy, May 2005), 29.
[117] Paul S. Canning, *A Definitive Work on Factors Impacting the Arming of Unmanned Vehicles*, NSWCDD TR-05/36 (Dahlgren, VA: Department of the Navy, May 2005), 29.
[118] Canning, "A Concept of Operations for Armed Autonomous Systems," 17 – 22.

not anti-personnel mines have helped establish the precedent of accepting automatic systems targeting other machines rather than people.[119] While Canning's targeting methodology involves some legal contortions, it may provide an intermediate step in developing fully autonomous lethal systems. Regardless of the engineering approach taken in designing an autonomous system capable of employing lethal force, the military must trust that the system will operate properly. Establishing this trust is the next area for investigation.

Establishing Trust and Preventing "Oops Moments"

On 12 October 2007 in Lohatla, South Africa, a South African National Defense Force (SANDF) Oerlikon 35mm Mk5 anti-aircraft weapon malfunctioned, tragically killing nine soldiers and seriously injured fourteen others.[120] Since the system has both manual and automatic operating modes, initial media reports speculated software errors might have caused the incident.[121] In fact, one witness said the officer who tried to shut the system down could not "because the computer gremlin had taken over."[122] While the investigating technical committee subsequently determined a mechanical failure caused the

---

[119] Canning, "Weaponized Unmanned Systems," 4 – 5.
[120] Noah Shachtman, "Robot Cannon Kills 9, Wounds 14," *Wired Danger Room: What's Next in National Security*, 18 October 2007, http://m.wired.com/dangerroom/2007/10/robot-cannon-ki/ (accessed 19 March 2012).
[121] Noah Shachtman, "Inside the Robo-Cannon Rampage," *Wired Danger Room: What's Next in National Security*, 19 October 2007, http://m.wired.com/dangerroom/2007/10/inside-the-robo/ (accessed 19 March 2012). Also see Shachtman, "Robot Cannon Kills 9, Wounds 14;" Tom Simonite, "'Robotic Rampage' Unlikely Reason for Deaths," *New Scientist*, 19 October 2007, http://www.newscientist.com/mobile/article/dn12812 (accessed 19 March 2012); and Leon Engelbrecht, "Did Software Kill Soldiers?" *IT Web*, 16 October 2007, http://www.itweb.co.za/index.php?option=com_content&view=article&id=6157&catid=96:defence-and-aerospace-technology (accessed 20 March 2012).
[122] Gavin Knight, "March of the Terminators: Robot Warriors are no Longer Sci-Fi but Reality. So What Happens When They Turn Their Guns on Us?" *Daily Mail (UK)*, 15 May 2009, http://www.dailymail.co.uk/sciencetech/article-1182910/March-terminators-Robot-warriors-longer-sci-fi-reality-So-happens-turn-guns-us.html (accessed 20 March 2012). Also see Singer, *Wired for War*, 196.

incident, the narrative that a software malfunction caused this incident continues to pervade media reporting of this incident.[123]

Peter Singer of the Brookings Institute terms incidents whereby computer malfunctions produce unexpected events as "oops moments."[124] Indeed, the event in Lohalta was not the first oops moment with autonomous systems. In October 1960, the ballistic missile early warning system reported the detection of a Soviet missile launch. As NATO prepared to respond in kind, technicians determined the system had detected the moon rather than missile launch.[125] Furthermore, in November 1969 technicians inadvertently loaded a test program with simulated missile launches into the actual missile warning system, causing US Strategic Command to scramble alert aircraft before realizing the mistake.[126] The narrative of oops moments highlights the importance of establishing the trust that autonomous systems will function properly and as designed.[127]

Former Air Force Chief Scientist Dr. Werner Dahm avers trust is essential for allowing the operation of autonomous systems. One of the current barriers to implementing higher levels of autonomy in unmanned systems is the lack of a means to accomplish the necessary verification and validation (V&V) of these systems.[128] Verification ensures "the

---

[123] "Were Lohatla Deaths an Accident?" *IOL News,* 25 January 2008, http://www.iol.co.za/news/south-africa/were-lohatla-deaths-an-accident-1.386941 (accessed 20 March 2012). For an example of the persistent narrative about a software malfuction, see Peter Finn, "A Future for Drones: Automated Killing," *The Washington Post*, 19 September 2011, http://www.washingtonpost.com/national/national-security/a-future-for-drones-automated-killing/2011/09/15/glQAVy9mgK_story_1.html (accessed 19 March 2012). Finn reports, "Malfunctions also are a problem: In South Africa in 2007, a semiautonomous cannon fatally shot nine friendly soldiers." Also see Singer, *Wired for War*, 196.
[124] Singer, *Wired for War*, 196.
[125] Singer, *Wired for War*, 197.
[126] Singer, *Wired for War*, 197.
[127] For research on the correlation between automation use and attitude toward automation, see Raja Parasuraman, "Humans and Automation: Use, Misuse, Disuse, Abuse," *Human Factors* 39, no. 2 (June 1997): 230 – 253.
[128] Werner J.A. Dahm, *Technology Horizons: A Vision for Air Force Science and Technology During 2010 – 2030*, AF/ST-TR-10-01-PR, (Washington, DC: Department of the Air Force, 15 May 2010), 42.

computer program of the computerized model and its implementation are correct," and validation substantiates that "a computerized model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model."[129] In essence, verification ascertains accuracy of model implementation, while validation assesses whether the model achieves its design goal.[130] Dahm argues, "Developing methods for establishing 'certification trust in autonomous systems' is the single greatest technical barrier that must be overcome to obtain the capability advantages that are achievable by increasing use of autonomous systems."[131] Developing the V&V capabilities to establish trust in autonomous systems will be essential for recognizing the full capabilities of these systems.

Software enables system autonomy.[132] Autonomy software, or autonomic software, has the functionality to make decisions, execute actions, and reason about both the system itself and its environment. A distinguishing characteristic of autonomy software is its ability to execute a "number of basic steps without human intervention in order to achieve a given goal."[133] As the number of steps the system can execute without human intervention increases, trust in the system typically decreases–a concern emerging from the increased potential for oops moments.[134] Since software failure has the potential to cause mission failure and possibly endanger human life, V&V is necessary to establish trust.[135] Validation of autonomy software poses additional challenges for

---

[129] Robert G. Sargent, "Verification and Validation of Simulation Models," *2005 IEEE Winter Simulation Conference* (Orlando, FL: IEEE, 4 – 7 December 2005), 1.
[130] United States Air Force Test Pilot School, "Flight Control System Ground Testing," *FQ 719A – Ground Testing Course Objectives*, (Edwards AFB, CA: Air Force Material Command, 1 March 2005), 1.
[131] Dahm, *Technology Horizons*, 42.
[132] Johann Scumann and Willem Visser, "Autonomy Software: V&V Challenges and Characteristics," *2006 IEEE Aerospace Conference* (Big Sky, MT: IEEE, 4 – 11 March 2006), 1.
[133] Schumann and Visser, "Autonomy Software," 2.
[134] Canning, *A Definitive Work on Factors Impacting the Arming of Unmanned Vehicles*, 7.
[135] Schumann and Visser, "Autonomy Software," 2.

establishing trust because of three issues: input space size and complexity; program logic complexity; and domain model and environment description size and complexity.[136]

V&V is necessary for all autonomous systems, so the V&V techniques and processes developed for other autonomous applications may provide methods to address the V&V challenges associated with lethal autonomous systems. Space exploration with robotic systems is an area that has witnessed an increase in autonomous operations. Communications lag resulting from the vast tyranny of distance associated with space exploration combined with the fiscal need to reduce operations cost associated with human operators in a control room are driving factors in the move to autonomous capabilities on National Aeronautics and Space Administration (NASA) spacecraft. Yet, due to the high costs of space missions and limited opportunities for space exploration, NASA mission managers are wary of incorporating new technologies on spacecraft unless engineers can provide strong safety and reliability guarantees.[137] NASA mission managers maintain the reluctance of Mrs. Weston to trust autonomous systems. V&V, therefore, becomes very important.[138] Many of the V&V processes NASA has developed for autonomous space systems will have applicability for lethal autonomous systems.

One autonomous system architecture NASA utilizes consists of a functional layer of robotic primitives and a decision layer with planning and execution functionality–an architecture structure that would have corresponding functions in a lethal autonomous system. The functional layer provides the system with a set of "standard, generic robot

---

[136] Schumann and Visser, "Autonomy Software," 3.
[137] Guillaume Brat and Ari Jonsson, "Challenges in Verification and Validation of Autonomous Systems for Space Exploration," *2005 IEEE International Joint Conference on Neural Networks* (Montreal, Quebec, Canada: IEEE, 31 July – 4 August 2005), 1.
[138] Brat and Jonsson, "Challenges in Verification and Validation of Autonomous Systems for Space Exploration," 1.

capabilities" that interface with the system hardware.[139] The decision layer enables the system to autonomously create and execute system operations to achieve specified operational tasks. Typical tasks of the decision layer include planning, scheduling, and monitoring system execution.[140] Each layer creates different V&V challenges. For example, the V&V of a planner deals with issues such as "correct manipulation of plans, domain model consistency checking, or correct resource utilization reasoning" while V&V of a functional component is more concerned with issues such as "run time errors, concurrency, and timing problems."[141] Similar challenges will emerge in the V&V of components within ethical decision-making architectures and when integrating functional components, such as targeting sensors, with the ethical decision-making system.

The wide range of environments for which planning software must develop solutions makes the task of verifying a planner extremely challenging. This verification challenge arises whether the planner is determining a route to take Martian soil samples or the route to attack a surface to air missile system. Despite the difficulties, techniques exist. Examining the planner's domain model with model checking techniques can verify certain model properties. Inconsistencies, ambiguities, and incompleteness are properties a model checker can look for. Another method is to actually verify the correct implementation of some basic planner capabilities. A plans developed by a planner are lists, or tree data structures. Therefore, part of the difficulty in checking the plan stems from the potential for unbounded growth of the plan.[142]

---

[139] Brat and Jonsson, "Challenges in Verification and Validation of Autonomous Systems for Space Exploration," 2.
[140] Brat and Jonsson, "Challenges in Verification and Validation of Autonomous Systems for Space Exploration," 3.
[141] Brat and Jonsson, "Challenges in Verification and Validation of Autonomous Systems for Space Exploration," 3.
[142] Brat and Jonsson, "Challenges in Verification and Validation of Autonomous Systems for Space Exploration," 4.

While verifying an executive layer is similar to verifying a decision layer, a problem emerges with assessing whether the executive layer interfaces correctly with planner and functional elements. To solve this conundrum, compositional verification provides a "divide-and-conquer technique that aims at taking advantage of the modular architecture of a system in order to decompose the hard (expensive) problem of system verification into manageable verification of its components."[143] The approach checks individual components against a local property. This methodology breaks the task into manageable components and avoids the "state-space explosion" arising from checking a system-level property on the entire system.[144]

NASA has also addressed the V&V challenge associated with learning systems, which would be present in a bottom-up or hybrid approach for a lethal autonomous system. NASA's work explored V&V methods for neural-network-based control systems.[145] For neural networks, NASA augmented V&V with "specifically tailored validation and dynamic monitoring tools."[146] The V&V of a neural-network-based system demonstrates the capability of the controller's model to reflect the actual system with sufficient accuracy. Consequently, the modeling error cannot exceed a certain threshold for the operating envelope. To accomplish this, engineers assign error bars to the neural network output. The error bar provides a prediction of the model output range with a 95 percent probability of the actual value falling within the specified range. Assuming a standard distribution of the errors,

---

[143] Brat and Jonsson, "Challenges in Verification and Validation of Autonomous Systems for Space Exploration," 5.
[144] Brat and Jonsson, "Challenges in Verification and Validation of Autonomous Systems for Space Exploration," 5.
[145] A neural network provides an architecture capable of evolving and responding to a changing environment.
[146] Pramod Gupta and Johann Schumann, "A Tool for Verification and Validation of Neural Network Based Adaptive Controllers for High Assurance Systems," *Eighth IEEE International Symposium on High Assurance Systems Engineering*, (Tampa, FL: IEEE, 25 – 26 March 2004), 1.

engineers can relate confidence intervals to the standard deviation of the network output. Given the history of training data and network parameters, engineers can calculate the error bars with a Bayesian approach. This technique enables engineers to establish the reliability of an adaptive learning system.[147]

While the V&V techniques and processes developed by NASA provide an approach to developing trust in autonomous systems, other paths also exist. Nuclear safety certification and certification of artificial intelligence systems in the medical field provide some additional paths for establishing trust.

Like NASA's V&V techniques, the nuclear safety certification program helps establish system trust by examining software.[148] It parses software with nuclear safety implications into three different categories, each of which has different levels of evaluation.[149] Table 5 presents the software categories for the nuclear safety certification program.

Table 5: Nuclear Safety Certification Program Software Categories[150]

Category	Software Purpose
I	Controls critical functions and/or designated as a critical component.
II	Controls critical functions but not designated as a critical component. Requires independent V&V.
III	Does not control critical functions but interfaces with hardware/software controlling critical functions.

Source: John S. Canning, A Definitive Work on Factors Impacting the Arming of Unmanned Vehicles

[147] Gupta and Johann Schumann, "A Tool for Verification and Validation of Neural Network Based Adaptive Controllers for High Assurance Systems," 2.
[148] For an analysis on the ability of complex organizations to manage hazardous technology, particularly nuclear weapons technology, see Scott D. Sagan, *The Limits of Safety: Organizations, Accidents, and Nuclear Weapons* (Princeton, NJ: Princeton University, 1993). Sagan evaluates the ability of organizations to prevent accidents in the management of nuclear weapons through the lens of both high reliability organizational theory and normal accidents theory.
[149] Canning, *A Definitive Work on Factors Impacting the Arming of Unmanned Vehicles*, 18.
[150] Canning, *A Definitive Work on Factors Impacting the Arming of Unmanned Vehicles*, 18.

By categorizing the software into different categories, engineers can focus V&V efforts on the most important functions. In other words, the categorization focuses scarce resources and time on the most important systems.

The nuclear safety certification program also provides a methodology for examining critical sequences. In doing so, it establishes probability standards for inadvertent events occurring in the sequence of employing a nuclear weapon. These critical functions have analogous functions in autonomous systems.[151] Table 6 lists the probabilities and consequences of these inadvertent events.

Table 6: Probability Standards for Inadvertent Events in Nuclear Weapons Employment Sequence[152]

Critical Function	Probability of Obtaining Nuclear Yield is Less Than	Circumstances
Authorization	None	Safety evaluations must consider authorization of the device as part of the command and control function.
Prearming	10^{-6} per delivery vehicle over system lifetime	Inadvertent transmission of prearm.
Arming	10^{-4} per prearmed weapon	Arming and fusing system failure resulting in arming after system has been prearmed but before launch.
Launching	10^{-7} per missile over the system's lifetime 10^{-12} per missile over the system's lifetime	Accidental propulsion system ignition Inadvertent programmed launch during fully assembled weapon system operation.
Releasing	10^{-6} per weapon station over the system's lifetime 10^{-3} per unlocking event	Inadvertent release or jettison of a bomb or missile when release system is locked Inadvertent release or jettison of a bomb or missile when release system is unlocked
Targeting	10^{-3} per missile 10^{-3} per delivery vehicle	Erroneous issuance of good guidance signal Inadvertent application of power or signals to warhead interface

Source: John S. Canning, A Definitive Work on Factors Impacting the Arming of Unmanned Vehicles, 18 – 19

[151] Canning, *A Definitive Work on Factors Impacting the Arming of Unmanned Vehicles*, 18.
[152] Canning, *A Definitive Work on Factors Impacting the Arming of Unmanned Vehicles*, 18.

These probabilities provide a potential yardstick for developing appropriate probabilities of critical events for lethal autonomous systems.[153] The general methodology provides value through its framework for assessing key steps in a critical sequence.

The medical community provides similar tools for analyzing the safety of intelligent systems utilized in safety critical functions. The use of intelligent systems to aid doctors initially faced resistance "given the potential consequences of adopting unproven methods."[154] Again, echoing Mrs. Weston, when presented intelligent systems to aid in medical decision-making, doctors asked themselves, "Why should I believe that this new technology is safe to use on my patients?"[155] To address this concern, the medical community adopted safety management techniques such as systematic hazard analysis and rigorous empirical testing. Since these methods do not prevent the emergence of unanticipated hazards, the medical community equipped artificial intelligent systems with an active safety management system to operate in parallel with the primary decision-making systems and to develop strategies for preventing and minimizing the consequences of unanticipated hazards.[156] For example, the OaSiS protocol management system for cancer incorporates these concepts. OaSiS safety principles include anticipation of adverse events, avoiding actions with the potential to exacerbate hazardous side effects, avoiding actions that could possibly diminish desirable effects, and reacting appropriately to hazards.[157] The safety management techniques and active safety management systems

---

[153] Canning, *A Definitive Work on Factors Impacting the Arming of Unmanned Vehicles*, 18.
[154] John Fox and Subrata Das, *Safe and Sound: Artificial Intelligence in Hazardous Applications* (Cambridge, MA: MIT Press, 2000), 132.
[155] John Fox and Subrata Das, *Safe and Sound: Artificial Intelligence in Hazardous Applications* (Cambridge, MA: MIT Press, 2000), 132.
[156] Fox and Das, *Safe and Sound*, 143.
[157] Fox and Das, *Safe and Sound*, 146 – 152.

provide additional methods for aiding in the establishment of trust in lethal autonomous systems.

While engineers have not yet solved the problem of establishing trust in autonomous systems, researchers have made considerable progress in this area. Techniques and processes developed by NASA for space exploration, nuclear certification, and the medical field provide potential paths for establishing trust in autonomous systems.

Conclusion

In the time between the World Wars, the ACTS air warfare strategists developed aerial bombardment plans and tactics despite lingering questions about the legality and morality of aerial bombardment. Similar questions about the morality and legality of lethal autonomous systems persist, yet the development of these systems is continuing apace.

Science fiction has provided fertile ground for thought about a robot ethic. Isaac Asimov's Three Laws of Robotics are probably the most famous. Asimov's Three Laws are simple and elegant in their formulation, yet his stories illustrate the hazards associated with attempting to implement these rules. In the end, Asimov's Three Laws provide a starting point for a discussion on robot ethics but do not offer a model for implementation.

The three basic approaches scientists are exploring as possibilities for implementing ethical reasoning in autonomous systems are top-down approaches, bottom-up approaches, and a hybrid approach, which combines both top-down and bottom-up techniques. Deontological top-down approaches have the potential to give rise to problems with slave morality, while utilitarianism top-down approaches lead to significant computational challenges. Particularism bottom-up approaches provide a dynamic morality, which can adapt to changing environments, but also create the potential to override control mechanisms. Hybrid approaches

combine methods and more accurately reflect the approach humans utilize. A hybrid approach appears to be the currently preferred method to approach the development of a moral reasoning capability for an autonomous system.

Ronald Arkin of the Georgia Institute of Technology has proposed an ethical-decision making architecture. Arkin's architecture consists of an ethical governor to constrain behavior, an ethical behavior control to ensure the application of ethical constraints, an ethical adaptor to improve system performance in the event of a system error, and a responsibility advisor to serves as the human-machine interface.

John Canning of the Naval Surface Warfare Division has proposed a targeting approach to blunt legal and moral concerns associated with lethal autonomous systems. Canning's approach would have autonomous systems target things rather than people, which is an approach with some historical precedent.

Trust is essential for the operation of autonomous systems. Verification and validation is a key process necessary to establish the requisite trust levels. As spacecraft have become more autonomous, NASA has begun developing verification and validation techniques for autonomous systems. As these verification and validation techniques under development mature, they will help alleviate some concerns with lethal autonomous systems. Techniques and processes for space nuclear certification and the medical field also provide potential paths for establishing trust in autonomous systems

Having explored potential methods for designing lethal autonomous systems, the focus now turns to examining how humans will interact with these systems and the implications of these interactions.

Chapter 3

Human Interaction with Autonomous Systems

> *The hope is that, in not too many years, human brains and computing machines will be coupled together very tightly, and that the resulting partnership will think as no human brain has ever thought and process data in a way not approached by the information-handling machines we know today.*
>
> — J.C.R. Licklider
> *Man-Computer Symbiosis*

As World War II began its dénouement in November 1944, Commanding General of the Army Air Forces Henry Hap Arnold looked to place research and development on "a sound and continuing basis" by having the scientific community develop a long-range plan for the Army Air Force.[1] General Arnold's vision for this program was to provide "well thought out, long-range thinking" to guarantee the security of the United States by providing "a guide for the next 10–20 year period."[2] In response to General Arnold's request, Dr. Theodore von Karman chaired a Scientific Advisory Board. The board's report asserted a new element of the technological character of war was "the decisive contribution of organized science to effective weapons."[3] It recommended an organizational construct to facilitate cooperation between the military and the scientific community. According to the board, the scientific process necessitated that "government authorities, military or civilian, should foster, but not dictate, basic research."[4] One of the areas the board recommended researching was "pilotless aircraft."[5] In fact, the

---

[1] Theodore von Karman, *Toward New Horizons: Science, Key to Air Supremacy* (Wright Field, Dayton, OH: Headquarters Air Material Command, May 1945), iii. Document is now declassified.
[2] von Karman, *Toward New Horizons*, iii.
[3] von Karman, *Toward New Horizons*, 1.
[4] von Karman, *Toward New Horizons*, 85.
[5] von Karman, *Toward New Horizons*, xxi.

board averred, "In the warfare of the future, pilotless airplanes…are bound to be of great importance."[6] As the board pondered future pilotless aircraft, it recognized the necessity of a mechanism for humans to interact with the machines by overriding or adjusting settings of the pilotless aircraft's automatic controls.[7]

Improving the performance of a concatenated human-machine system is a complex, challenging problem requiring the exploitation of strengths and mitigation of weaknesses.[8] Humans excel at certain tasks, while machines excel at others.[9] For example, automatic pilots, systems known today as autopilots, developed in the 1930s were able to maintain straight and level flight better than human pilots could.[10] In fact, the flight manual for the World War II era B-17 noted the autopilot "detects flight deviations the instant they occur, and just as instantaneously operates the controls to correct the deviations."[11] Conversely, human pilots performed better during unpredictable situations and those situations requiring rapid maneuvers, such as aerial combat.[12] Thus, leveraging strengths and weaknesses improves the overall human-machine system performance. By the beginning of the twenty-first century, however, technological advancements had improved machine performance to point where Tony Tether, Director of the Defense Advanced Research Projects Agency (DARPA), warned a significant

---

[6] W.H. Pickering, "Automatic Control of Flight," in *Guided Missiles and Pilotless Aircraft: A Report of the AAF Scientific Advisory Group*, eds. H.L. Dryden, W.H. Pickering, H.S. Tsien, and G.B. Schubauer, (Wright Field, Dayton, OH: Headquarters Air Material Command, May 1946), 17. Document is now declassified.
[7] W.H. Pickering, "Automatic Control of Flight," 17.
[8] Colonel Timothy P. Schultz, "Redefining Flight: How the Predecessors of the Modern United States Air Force Transformed the Relationship Between Airmen and Aircraft," (PhD diss., Duke University, 2007), 151.
[9] Paul M. Fitts, *Human Engineering for an Effective Air-Navigation and Traffic-Control System* (Washington, DC: National Research Council, March 1951), 6 – 8.
[10] David A. Mindell, *Between Human and Machine: Feedback, Control, and Computing before Cybernetics* (Baltimore, MD: Johns Hopkins University, 2002), 138.
[11] "Pilot Training Manual for the B-17 Flying Fortress," HQ Army Air Forces, Office of Flying Safety, 183. Special manuscript series 603, box 3, Clark Special Collections Branch, USAF Academy Library, CO. Quoted in Schultz, "Redefining Flight," 152.
[12] Mindell, *Between Human and Machine*, 138.

challenge confronting the US Department of Defense was "preventing human performance from becoming the weakest link the on the future battlefield."[13]

Developing the appropriate interface between human and machine is also a daunting and complex problem. In *Man-Computer Symbiosis*, computer scientist J.C.R. Licklider postulates a primary aim of the cooperative interaction between humans and computers is to facilitate decision-making and to control complex situations. Through this symbiosis, Licklider argues, complex problems "would be easier to solve, and they could be solved faster, through an intuitively guided trial-and-error procedure in which the computer cooperated, turning up flaws in the reasoning or revealing unexpected turns in the solution."[14] A symbiotic relationship between human and computer would take advantage of the positive characteristics of both to produce more effective and efficient capabilities.[15] Yet despite the promises of human-computer symbiosis, achieving Licklider's vision has proven to be difficult.

In his 1985 book *Command in War*, Military historian Martin van Creveld explores the evolution of command and control systems.[16] He notes the interaction between humans and machines raises many salient questions.

> The significance of the technological revolution for the problems of command is even clearer when it is seen that the last three decades have produced, for the first time in history, artificial devices capable of reproducing or amplifying the functions not merely of man's limbs and sensory organs but, to a growing extent, those of his brain as well. This has given birth to a host of questions for which little or no precedent existed. Which are the strong points of

---

[13] Tony Tether, Director, Defense Advanced Research Projects Agency (statement before the Subcommittee on Military Research and Development, Committee on Armed Services, Washington, DC, 26 June 2001). The statement is available at www.darpa.mil/WorkArea/DownloadAsset.aspx?id=1781 (accessed 3 May 2012).
[14] J.C.R. Licklider, "Man-Computer Symbiosis," *IRE Transactions on Human Factors in Electronics* 1, March 1960, 5.
[15] Licklider, "Man-Computer Symbiosis," 6.
[16] Martin van Creveld, *Command in War* (Cambridge, MA: Harvard University, 1985), 1.

> man, and which are those of the new machines? How, in consequence, should the burden of work be divided among them? How should communication ("interface") between man and machine, as well as among the machines themselves, be organized?[17]

Solutions to these questions raised by van Creveld look to utilize strengths and minimize weaknesses of both humans and machines to produce a more effective and efficient system.

Technological advances have not only created a symbiotic relationship, they have blurred the boundaries between humans and machines.[18] This blurring of boundaries has been a long, on-going process. During World War I, the hit rate for the British Navy at the Battle of Jutland was an embarrassingly low three percent. The one exception to this dismal performance arose from the British ship armed with a mechanized calculating system. This exception piqued the US Navy's interest. The American inventor-engineer Hannibal Ford designed the Mark 1 Ford Rangekeeper to solve the multiple calculations necessary to direct naval gun fire at the enemy ship's predicted location. Ford's Rangekeeper required gunnery officers to input variables into the system via hand cranks.[19] Recognizing the importance of both humans and machines in solving the naval gunnery problem (and in a stroke of marketing genius), Ford sold his device to the US Navy "not as a replacement for the skilled officers but rather as an aid, an instrument that would both require and enhance the prestige of their mathematical skill."[20]

The blurring trend continued between the World Wars as engineers developed aircraft bombsights based on automatic pilots and fire control systems for anti-aircraft artillery.[21] The exploits of the American aviator

---

[17] van Creveld, *Command in War*, 2 – 3.
[18] Mindell, *Between Human and Machine*, 2.
[19] Mindell, *Between Human and Machine*, 20 – 21.
[20] Mindell, *Between Human and Machine*, 21.
[21] Mindell, *Between Human and Machine*, 43, 82.

Wiley Post highlight the blurring boundary between human and machine. In 1931, Post the set a world record for the fastest round-the-world flight in his Vega aircraft named the *Winnie Mae*. Two years later, he repeated the feat, this time flying solo. A Sperry A-2 automatic pilot was essential for Post to accomplish this solo flight.[22] Noting the importance of the A-2 in Post's feat, the *New York Times* predicted, "The days when human skill and almost bird-like sense of direction enabled a flier to hold his course for long hours through a starless night or over a fog are over … Commercial flying in the future will be automatic."[23]

Work on automation technology such as autopilots and anti-aircraft artillery fire control systems gave rise to the field of inquiry Massachusetts Institute of Technology (MIT) mathematics professor Norbert Wiener termed cybernetics. Wiener defined cybernetics as "the entire field of control and communication theory" and derived the term from the Greek word for steersman.[24] Cybernetic systems sense and react to external inputs. Information feedback is a critical component of this process. A cybernetic system converts information into action. This action subsequently generates information for future action. As information feeds back into the system, it continues to self-regulate.[25] Two simple examples of this self-regulation process are a thermostat and a steam engine governor.[26] As the line between human and machine blurred, human operators became a "manual servomechanism," or a machine within the larger machine, holding the integrated system together.[27]

While automation and autonomy have many advantages, they also have a dark side. As the line has blurred, humans have slowly ceded

---

[22] Mindell, *Between Human and Machine*, 78 – 80.
[23] Quoted in Mindell, *Between Human and Machine*, 80.
[24] Norbert Wiener, *Cybernetics, or Control and Communication in the Animal and the Machine*, 2d ed. (New York: MIT Press, 1961), 19.
[25] Schultz, "Redefining Flight," 147.
[26] Wiener, *Cybernetics*, 96 – 97.
[27] Mindell, *Between Human and Machine*, 92 – 93, 98.

aspects of lethal decision-making authority to machines. Modern military systems such as the Aegis combat system and the Patriot surface-to-air missile have the capability to make lethal decisions autonomously. Thus, while critics of autonomous systems are loathe to allow these systems to make lethal decisions, the US military has already embarked along this path.[28] This increasing autonomy has consequences. Research on mishaps involving automation technology has demonstrated "excessive trust can lead operators to rely uncritically on automation without recognizing its limitations or fail to monitor the automation's behavior."[29] The Aegis and Patriot missile both allow the operator to override the system's judgment, yet humans are often reluctant to do so, producing tragic consequences.[30]

The design of a system influences the nature of human-machine interactions. Human factors engineering, previously known as human engineering, is a "special branch of applied psychology that deals with the relations between men and machines…to find out what the capabilities and limitations of human beings are in using various kinds of equipment."[31] It helps answer the question, "What should men do and what should machines do?"[32] Thus, as engineers develop lethal autonomous systems, human factors engineering will have a significant role because how humans interact with lethal autonomous systems will have ethical and moral implications.

This chapter will examine insights from human factors engineering on the interaction between humans and machines, how this interaction affects the psychology of killing, the influence of automation bias in the

---

[28] Peter W. Singer, *Wired for War: The Robotics Revolution and Conflict in the 21st Century* (New York: Penguin, 2009), 124.
[29] Raja Parasuraman, "Humans and Automation: Use, Misuse, Disuse, Abuse," *Human Factors* 39, no. 2 (June 1997), 239.
[30] Singer, *Wired for War*, 125.
[31] Fitts, *Human Engineering for an Effective Air-Navigation and Traffic-Control System*, 2.
[32] Fitts, *Human Engineering for an Effective Air-Navigation and Traffic-Control System*, 2.

interaction between humans and machines, and the responsibilities of those developing and designing autonomous systems.

Insights from Human Factors Engineering

In the first half of the twentieth century, engineers often overlooked the human component when designing machines. For example, in 1921 the Martin Company neglected to equip its MB-2 bomber with a windshield to protect the pilot from environmental conditions such as windblast and rain. The Engineering Division of the Army Air Service had to recommend that Martin correct this oversight.[33]

Safety was frequently an afterthought in design. In March 1910, Lieutenant Benjamin Foulois's crash in the Army's first military airplane highlighted the necessity of having a safety belt. Following the crash, he wrote, "The two truss wires in front of the pilot's seat saved me from being thrown completely out of the airplane…. Thereafter, I used a four-foot trunk strap with which I lashed myself to the pilot's seat."[34] In another flagrant safety oversight, the N-3A gunsight of World War II era P-39 and P-40 aircraft protruded eleven inches from the instrument panel, placing the device in an ideal position to inflict massive head injuries to the pilot in the event of an abrupt stop.[35]

The prevailing engineering approach was to develop machines with the mindset that humans would adapt and adjust to machines rather than designing the machines to compensate for human limitations. The most egregious manifestations of this design mindset, however, led to the development of "mechanical monstrosities" that prevented effective task accomplishment by inhibiting the integration of human and machine.[36]

---

[33] Schultz, "Redefining Flight," 158 – 159.
[34] Major General Benjamin D. Foulois, USAF (Ret), "Early Flying Experience in Army Airplane No. 1 (1909-1910-1911)," unpublished article, 1960. Manuscript series 17, box 6 Clark Special Collections Branch, USAF Academy Library, CO. Quoted in Schultz, "Redefining Flight," 156.
[35] Schultz, "Redefining Flight," 161.
[36] Fitts, *Human Engineering for an Effective Air-Navigation and Traffic-Control System*, iv.

By the 1930s, however, the principles of human factors engineering began to emerge. The 1934 Army Air Corps *Handbook of Instructions for Airplane Designers* directed the arrangement of aircraft instruments and controls in the cockpit for "maximum comfort and convenience."[37] While the mindset shift began before World War II, the flood of military requirements emerging from wartime exigencies precipitated increasing acceptance of the principle that "machines should be made for men; not men forcibly adapted to machines."[38]

The 1951 report *Human Engineering for an Effective Air-Navigation and Traffic-Control System* led by Dr. Paul M. Fitts of The Ohio State University was a pioneering work that looked to resolve differences between design engineers and human factors engineers in developing an effective human-machine interface. The report provided a long-range program of psychological research for researchers to conduct concurrently with equipment and system development.[39] Many of the human factors engineering principles distilled in the Fitts Report retain their relevance today.[40] With proper human factors engineering, it becomes possible to "redesign many pieces of equipment so that the 'human errors' are greatly reduced or even eliminated."[41]

One of the principle questions the Fitts Report committee members investigated was the role assigned to humans in control systems involving automation. This area of investigation retains its salience since autonomous systems differ primarily in the number of steps

---

[37] *Handbook of Instructions for Airplane Designers*, 7th ed., vol. 1 (Wright Field, OH: U.S. Army Air Corps, 1934), 317. History Office, Air Force Material Command (AFMC), Wright Patterson Air Force Base, Ohio. Quoted in Schultz, "Redefining Flight," 161.
[38] Fitts, *Human Engineering for an Effective Air-Navigation and Traffic-Control System*, iv.
[39] Fitts, *Human Engineering for an Effective Air-Navigation and Traffic-Control System*, iv.
[40] John Reising, "Fitts' Principles Still Applicable: Computer Monitoring of Fighter Aircraft Emergencies," *Aviation Space and Environment Medicine* 53, no. 11 (Nov 1982): 1080 – 1084. The article is available on line at http://www.faa.gov/library/online_libraries/aerospace_medicine_/sd/media/reising_j.pdf (accessed 25 March 2012).
[41] Committee on Automation in Combat Aircraft, *Automation in Combat Aircraft* (Washington, DC: National Research Council, 1982), V-100.

accomplished without human involvement. The committee postulated four possibilities: fully automatic control, automatic control with human monitoring; semi-automatic control supplemented by human performance of critical functions; and primary control by human operators assisted by effective data-analysis, data-transmission, and data-display equipment.[42] Figure 10 illustrates these roles along a spectrum of autonomy.

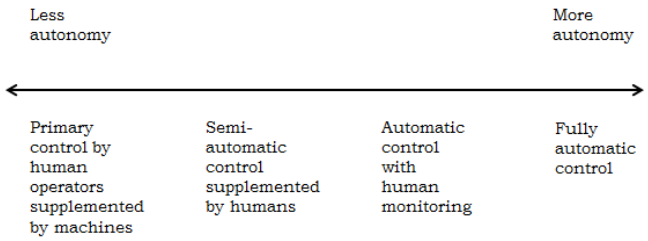

Figure 10: Roles Assigned to Humans in Control Systems with Automation

Source: Adapted from Paul Fitts, Human Engineering for an Effective Air-Navigation and Traffic-Control System, 5

The level of human involvement in such systems affects the human engineering and design of the system. For example, a system designed for automatic control with human monitoring would center on the capabilities of the human as a monitor, trouble-shooter, or emergency controller while facilitating those specific functions.[43]

The committee examined the relative strengths of humans and machines. Human strengths included sensory functions, perceptual abilities, flexibility, judgment and selective recall, and reasoning.[44] The committee asserted humans are able to size up complex situations quickly, particularly when the systems present data in a manner to maximize human perceptual capability through pictorial or patterned

[42] Fitts, *Human Engineering for an Effective Air-Navigation and Traffic-Control System*, 5.
[43] Fitts, *Human Engineering for an Effective Air-Navigation and Traffic-Control System*, 5.
[44] Fitts, *Human Engineering for an Effective Air-Navigation and Traffic-Control System*, 7 – 8.

displays.[45] The committee also reasoned that humans ought to perform important roles in systems requiring flexibility. Machines are inflexible. They are limited to their programming, which limits the solution set, whereas humans can develop multiple solutions to the same problem. Therefore, in instances where engineers could not reduce operations to "logical, pre-set procedures," humans are necessary to exercise judgment.[46]

This flexibility helps prevent complete breakdowns during emergencies, which can be highly contextual and extend beyond pre-set procedures. For example, consider the resolution of the problem facing the crew of Apollo 13 with increasing cabin levels of carbon dioxide. Mission control engineers developed an adapter for the square command module canister that the crew used in place of the cylindrical air scrubbers in the lunar module. Flight Director Gene Kranz described the solution as a "rather bizarre but functional contraption."[47] While a machine could have alerted the crew to change filters, it is highly unlikely a machine with Apollo era technology could have developed such a unique solution.[48] In short, humans possess "remarkable powers that cannot yet be duplicated by machines, especially abilities needed to deal with changing situations and unforeseen problems."[49]

On the other side of the equation, machine strengths included speed and power, performing routine work, computational ability, short-term storage of information, and the ability to perform simultaneous activities.[50] While machines are not infallible, they can "carry out specific

---

[45] Fitts, *Human Engineering for an Effective Air-Navigation and Traffic-Control System*, 7.
[46] Fitts, *Human Engineering for an Effective Air-Navigation and Traffic-Control System*, 7.
[47] Gene Kranz, *Failure is not an Option: Mission Control from Mercury to Apollo 13 and Beyond* (New York: Simon & Schuster, 2000), 328.
[48] It is possible a modern intelligent system, such as IBM's Watson, could have developed a solution similar to the adaptor created by mission control engineers. For Watson to offer such a solution, however, it would have needed an analogous case available in its search database.
[49] Fitts, *Human Engineering for an Effective Air-Navigation and Traffic-Control System*, 8.
[50] Fitts, *Human Engineering for an Effective Air-Navigation and Traffic-Control System*, 8.

functions with fewer errors than would be made by humans," produce quicker and more uniform responses, and do not become bored and inattentive.[51] Based on these observations on the relative strengths of humans and machines, the committee concluded human tasks should "provide activity," be "intrinsically interesting," and "as a rule machines should monitor men."[52]

Role allocation between humans and machines becomes "one of the most critical aspects of any integrated sociotechnical system design with significant embedded autonomy."[53] Technical capabilities must inform decisions engineers make in determining which systems they automate and the extent to which they apply automation.[54] Applying automation or autonomy indiscriminately without regard to the roles and responsibilities of the operator can adversely affect the system's overall performance.[55] Design engineers must make this allocation decision within the context of the human operator's capabilities and limitations, such as perceptual skills, information processing capabilities, and physiological restrictions.[56] After deciding where to apply automation, engineers must evaluate the impact of the role allocation on human performance.[57]

---

[51] Fitts, *Human Engineering for an Effective Air-Navigation and Traffic-Control System*, 6.
[52] Fitts, *Human Engineering for an Effective Air-Navigation and Traffic-Control System*, 6.
[53] Mary L. Cummings and K.M. Thornburg, "Paying Attention to the Man Behind the Curtain," http://www.web.mit.edu/aeroastro/labs/halab/papers/Final_Curtain.pdf (accessed 21 March 2012), 4. This paper was published in *IEEE Pervasive Computing* 10 (Jan – Mar 2011), 58 – 62.
[54] Raja Parasuraman, Thomas B. Sheridan, and Christopher D. Wickens, "A Model for Types and Levels of Human Interaction with Automation," *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans* 30, no. 3 (May 2000), 286.
[55] Parasuraman, "Humans and Automation," 247.
[56] USAF Test Pilot School, "Technology and Automation," *Systems Phase Text Book Chapter 3 – Human Factors* (Edwards AFB, CA: Air Force Material Command, July 2002), 3-38.
[57] Parasuraman, Sheridan, and Wickens, "A Model for Types and Levels of Human Interaction with Automation," 289 – 290.

Automation provides a mechanism to decrease excessive workload, reduce errors, improve performance, and add capabilities.[58] In fact, the lure of reducing errors in the application of lethal force is a strong motivating factor for developing lethal autonomous systems.[59] Some typical sources of operator workload include perceptual saturation, concurrently performed tasks, time-line compression, operator bandwidth limitations, and small-scale, routine operations.[60] While these workload sources have the potential to adversely affect operator performance, properly applying automation can provide a means to mitigate these detractors.[61] Yet despite these potential benefits, automation is not a panacea. Rather, automation "changes the nature of the work that humans do, often in ways unintended and unanticipated by the designers of automation."[62] In some instances, poor human factors engineering in the design of the human-machine interface can drive human errors or increase the system's overall inefficiency.[63]

While automation processes are useful for quickly and precisely executing repeatable computations in applications such as complex optimization problems like solving power distribution issues in smart electrical grid operations, they can also be inflexible and unable to adapt to changing situations. This inflexibility and inability to adapt arises because these applications can only account for variables identified

---

[58] Committee on Automation in Combat Aircraft, *Automation in Combat Aircraft* (Washington, DC: National Research Council, 1982), V-17

[59] Ronald C. Arkin, *Governing Lethal Behavior in Autonomous Robots* (New York: CRC Press, 2009), 29 – 36.

[60] USAF Test Pilot School, "Technology and Automation," 3-38. Operator bandwidth limitations arise from temporal limitations of humans in accomplishing tasks. For example, the frequency of inputs required to control the aerodynamically unstable F-16 exceeds the human ability to make inputs. Thus, F-16 flight control computers augment human capabilities and make automatic inputs to control the aircraft.

[61] USAF Test Pilot School, "Technology and Automation," 3-38.

[62] Raja Parasuraman, "Humans and Automation: Use, Misuse, Disuse, Abuse," *Human Factors* 39, no. 2 (June 1997), 231.

[63] Anthony P. Tvaryanas, "Human Systems Integration in Remotely Piloted Aircraft Operations," *Aviation, Space, and Environmental Medicine* 77, no. 12 (December 2006), 1282.

during the design stage.[64] In contrast, humans solve problems through knowledge-based reasoning, whereby they utilize abilities to improvise, learn, and reason inductively to solve problems.

Thus, in complex control systems, automation manages most problem solving and system management. Humans augment the system by responding to dynamic and unexpected events.[65] For example, automated systems handle most of the energy distribution tasks with power grids, but humans remain in the loop to respond to unexpected situations.[66] Yet having humans in the loop to back-up autonomous systems cannot always mitigate poor system design. For example, the August 2003 blackout in the northeastern United States occurred in part because the design of the power management system did not provide a mechanism to alert control room operators of a degraded operating capability. Therefore, since the system's alarms failed, the operators did not realize the system was degrading and could not take the appropriate steps to avert the impending disaster.[67] The northeastern blackout demonstrates that, if designers expect human operators to monitor machines, then human factors engineering must provide a way for the operators to intervene.

While maintaining humans in the loop of control systems to intervene with the occurrence of an unexpected event is an established concept, researchers are still working to ascertain how well human operators can assist automatic systems with optimizing overall system performance.[68] Military research in the area of managing decentralized networks of unmanned vehicles is providing insights into how designers

---

[64] Cummings, "Paying Attention to the Man Behind the Curtain," 4.
[65] Cummings, "Paying Attention to the Man Behind the Curtain," 4.
[66] Cummings, "Paying Attention to the Man Behind the Curtain," 4 – 5.
[67] U.S.-Canada Power System Outage Task Force, *Final Report on the August 14, 2003 Blackout in the United States and Canada: Causes and Recommendations* (Washington, DC: Department of Energy, April 2004), 51 – 52. The report is available on-line at https://reports.energy.gov/ (accessed 26 March 2012).
[68] Cummings, "Paying Attention to the Man Behind the Curtain," 5.

should enable humans to collaborate in this process.[69] In one such experiment, a single human managed a team of unmanned systems tasked with locating and identifying targets. After the operator set priorities, an autonomous planner developed a task distribution list for the network. The operator subsequently accepted, rejected, or modified the plan.[70] Allowing the human operator to occasionally modify the solution produced a search pattern that located and identified more targets than did the solution generated by the planner.[71]

The human operator becomes critical in such a system because of its inherent uncertainty. Theoretically, the autonomous system planners generate an optimal solution. In reality, however, occasional human judgment can actually improve performance.[72] It is not uncommon in command-and-control scenarios for automated and human decision makers to differ on what is a better solution. Satisficing behavior and assessments of qualitative variables helps explain this difference.[73] Yet too much human intervention can hinder system performance. The key becomes developing a "robust range of helpful human interaction."[74]

Despite the perceived ability of humans to outperform machines in managing complex, unexpected occurrences, the human track record in this area is not particularly strong, as illustrated by the 1986 Chernobyl disaster and the 2003 blackout in the northeastern United States. Humans, too, face difficulties in understanding and managing complex

---

[69] Cummings, "Paying Attention to the Man Behind the Curtain," 6.
[70] Mary L. Cummings, Andrew Clare, and Christin Hart, "The Role of Human-Automation Consensus in Multiple Unmanned Vehicle Scheduling," *Human Factors: The Journal of the Human Factors and Ergonomics Society* 52, no. 1 (February 2010), 18 – 19.
[71] Cummings, "Paying Attention to the Man Behind the Curtain," 6.
[72] Cummings, "Paying Attention to the Man Behind the Curtain," 6.
[73] Cummings, Clare, and Hart, "The Role of Human-Automation Consensus in Multiple Unmanned Vehicle Scheduling," 18.
[74] Cummings, "Paying Attention to the Man Behind the Curtain," 7.

problems.[75] Poor human factors engineering can exacerbate the situation. In the Three Mile Island accident and the 2003 Northeast blackout, a significant problem was "the lack of explicit design to support rapid data aggregation and information visualization to support supervisors' time-pressured decision making."[76]

Critics of autonomous systems would leave lethal decision making solely in the hands of humans. Yet humans are capable of making tragic mistakes without the assistance of autonomous systems, particularly in war. Two recent fratricide incidents illustrate human fallibility in warfare–one in An Nasiriyah, Iraq and the other in Pashmul, Afghanistan.

On March 23, 2003, a flight of two A-10As mistakenly attacked a detachment of Marines near An Nasiriyah, Iraq.[77] Eighteen Marines died during the battle, though not all by friendly fire.[78] Unaware that elements of the battalion had crossed the Saddam Canal Bridge, a company commander cleared the A-10s to engage targets north of the bridge.[79] The resulting attack from the A-10s killed several Marines and destroyed multiple US vehicles.[80]

A similar friendly-fire incident occurred on September 4, 2006, when an A-10A mistakenly attacked a Canadian position near Pashmul, Afghanistan during Operation MEDUSA, killing one Canadian soldier.[81] The A-10 pilot mistakenly identified a trash fire as the target area and

---

[75] Dietrich Döner, *The Logic of Failure – Why Things Go Wrong and What We Can Do to Make Them Right*, trans. Rita and Robert Kimber, (New York: Metropolitan Books, 1996), 37.
[76] Cummings, "Paying Attention to the Man Behind the Curtain," 1.
[77] United States Central Command, *Investigation of Suspected Friendly Fire Incident Near An Nasiriayah, Iraq, 23 March 2003* (Tampa, FL: US Central Command, 6 March 2004), 2. The document is declassified.
[78] United States Central Command, *Investigation of Suspected Friendly Fire Incident Near An Nasiriayah, Iraq, 23 March 2003*, 2 – 3.
[79] United States Central Command, *Investigation of Suspected Friendly Fire Incident Near An Nasiriayah, Iraq, 23 March 2003*, 1 – 3.
[80] United States Central Command, *Investigation of Suspected Friendly Fire Incident Near An Nasiriayah, Iraq, 23 March 2003*, 1 – 3.
[81] Seth G. Jones, *In the Graveyard of Empires: America's War in Afghanistan* (New York: Norton, 2010), 215 – 216.

strafed the position.[82] Tragically, the pilot had situational awareness cues available that should have indicated that he was attacking the wrong target, leading the Canadian Expeditionary Forces Board of Inquiry to conclude the incident was preventable.[83]

In both of these instances, preventable human mistakes led to fratricide incidents. Furthermore, technology could have augmented human decision-making in both these incidents to provide greater battlefield awareness, possibly preventing both incidents. For example, in both instances, Blue Force tracking technology such as the information displayed by the situational awareness data link (SADL) could have increased the pilot's awareness of friendly location, possibly preventing these incidents.[84] In Ronald Arkin's proposed architecture for an autonomous lethal system, the system would weigh the location of friendly forces in the decision to apply lethal force. The proportionality algorithm of the ethical governor would specifically restrict the application of force if it could result in fratricide.[85] While technology can aid humans in applying lethal force, it can also impact a person's willingness to apply lethal force.

On the Psychology of Killing

In the early 1960s, Yale psychology professor Stanley Milgram conducted experiments to ascertain why humans obey commands to inflict harm on others. Milgram's studies disturbingly demonstrated that

---

[82] Canadian Expeditionary Forces, *Board of Inquiry Minutes of Proceedings A-10A Friendly Fire Incident 4 September 2006, Panjwayi District, Afghanistan*, (Ottawa, Canada: Department of National Defense, 13 July 2007), 14 – 15. The document is declassified. The report is available on-line at http://www.forces.gc.ca/site/focus/opmedusa/A10_BOI_Report_e.pdf (accessed 26 April 2012).
[83] Canadian Expeditionary Forces, *Board of Inquiry Minutes of Proceedings A-10A Friendly Fire Incident 4 September 2006, Panjwayi District, Afghanistan*, 4.
[84] Maj Sean P. Larkin, "Air-to-Ground Fratricide Reduction Technology: An Analysis," (master's thesis, Marine Corps Command and Staff College, 2005), 1 – 3.
[85] Arkin, *Governing Lethal Behavior in Autonomous Robots*, 187. Figures 12.6 and 12.7 on page 186 provide an illustration of how the location of friendly combatants would affect the system's choice of weapon when applying lethal force.

people would "obey authority to a greater extent" than expected.[86] To assess the subject's willingness to obey a command to inflict harm, Milgram's research team directed the subject to administer increasingly stronger electric shocks to a victim. The victim was part of the research team and did not actually receive a shock but acted as though he did.[87]

In a variation of the experiment, Milgram's researchers altered the psychological distance between the victim and the subject. In the first instance, the subject could not see or hear the victim who was located in another room.[88] In the second design, the subject could hear but not see the victim. For the third design, the victim was in the same rooms as the subject. The final experimental design required the subject to physically place the victim's hand on a plate to receive the electric shock. The percentage of subjects who disobeyed commands to continue administering the electric shock increased from 34 percent in the remote condition to 70 percent in the touch-proximity condition.[89]

Milgram postulated that empathic cues were one of the mechanisms that explained this increasing disobedience. As the psychological distance from the victim increases, the victim's suffering is "an abstract, remote quality for the subject."[90] The person inflicting pain and suffering on the victim is aware of the punishment inflicted on the victim, but "only in a conceptual sense…the fact is apprehended, but not felt."[91] Milgram believed the same detachment occurred during wartime. For example, he argued that a bombardier would reasonably know "that

---

[86] Stanley Milgram, "Some Conditions of Obedience and Disobedience to Authority," *Human Relations* 18, no. 1 (February 1965): 61.
[87] Milgram, "Some Conditions of Obedience and Disobedience to Authority," 61.
[88] With this set-up, the victim banged on the wall in protest when the subject administered the 300-volt shock. After receiving the 315-volt shock, the victim provided no more feedback to the subject.
[89] Milgram, "Some Conditions of Obedience and Disobedience to Authority," 61 – 62. The percentages of disobedience were 34 percent in the remote condition, 37.5 percent when the subject received audible feedback from the victim, 60 percent when the victim and subject were in the same room, and 70 percent when the subject had to physically place the victim's hand on the shock plate.
[90] Milgram, "Some Conditions of Obedience and Disobedience to Authority," 63.
[91] Milgram, "Some Conditions of Obedience and Disobedience to Authority," 63.

his weapons will inflict suffering and death, yet this knowledge is divested of affect, and does not move him to a felt, emotional response to the suffering resulting from his actions."[92]

Milgram also hypothesized that remoteness enabled denial and a narrowing of the cognitive field. Distance enabled the subject to narrow his or her cognitive field, thus placing the victim out of mind. Since the subject disassociated his or her act of pressing a lever to administer a shock from the victim's suffering, the subject's action is no longer "relevant to moral judgment."[93] With proximity, this coping mechanism becomes more difficult.

Milgram further posited that proximity enabled reciprocal fields, whereby not only was the subject able to see the victim, but the victim was able to see the subject. This proximity allowed the victim to scrutinize the subject's actions. Knowing another person is observing one's actions can generate "shame, or guilt, which may then serve to curtail the action."[94] Face-to-face confrontation is psychologically discomforting, and therefore, it inhibits some actions.

Lieutenant Colonel Dave Grossman, a retired Army Ranger and psychologist, applies Milgram's observations to explain the reluctance of soldiers to kill on the battlefield in his 2009 work *On Killing*. From a historical perspective, soldiers on the battlefield have been reluctant to fire their weapons because of a powerful and innate resistance to killing.[95] Yet the firing rate among soldiers increased from 15 percent of soldiers firing their weapons in World War II to 55 percent in the Korean War and up to 90 percent in the Vietnam War.[96] Grossman argues this trend has occurred because militaries have developed modern combat training techniques "to condition soldiers to overcome their resistance to

---

[92] Milgram, "Some Conditions of Obedience and Disobedience to Authority," 63.
[93] Milgram, "Some Conditions of Obedience and Disobedience to Authority," 63.
[94] Milgram, "Some Conditions of Obedience and Disobedience to Authority," 64.
[95] Dave Grosman, *On Killing: The Psychological Cost of Learning to Kill in War and Society* (New York: Back Bay Books, 2009), xxxi.
[96] Grosman, *On Killing*, 26.

killing."[97] Desensitization and denial defense mechanisms also combined with conditioning to create the conditions for the tremendous increase in firing rates among soldiers.[98]

The psychological burden of killing is extremely high. Impersonalizing the act of killing another human provides a mechanism for a soldier to cope with the associated guilt.[99] The language many soldiers use to describe killing on the battlefield reflects this coping mechanism. Soldiers describe enemy combatants as being "knocked over, wasted, greased, taken out, and mopped up" rather than killed.[100] Soldiers further impersonalize the act of killing by denying the enemy's humanity by utilizing derogatory names for the enemy such as "Kraut, Jap, Reb, Yank …"[101]

John Canning's notion of targeting the archer's bow or arrows rather than the archer reflects this impersonalization of the act of killing.[102] During World War II, a view of the enemy arose from the mobilization of the scientific community in support of the war effort. For these scientists, the enemy emerged as "a cold-blooded, machinelike opponent … a mechanized Enemy Other."[103] Within the scientific community, operations research, game theory, and cybernetics engaged this version of the enemy. Operations research focused on improving the efficiency of wartime operations, while game theory provided a mechanism for strategists to analyze the strategic dilemmas.[104] Through cybernetics, the enemy operator became "so merged with machinery that

---

[97] Grosman, *On Killing*, xxxi.
[98] Grosman, *On Killing*, 253.
[99] Grosman, *On Killing*, 86 – 90.
[100] Grosman, *On Killing*, 91.
[101] Grosman, *On Killing*, 91.
[102] John S. Canning, "Weaponized Unmanned Systems: A Transformational Warfighting Opportunity, Government Roles in Making it Happen," *Engineering the Total Ship Symposium* (Falls Church, VA: American Society of Naval Engineers, 23 – 25 September 2008), 4.
[103] Peter Galison, "The Ontology of the Enemy: Norbert Wiener and the Cybernetic Vision," *Critical Inquiry* 21, no. 1 (Autumn 1994), 231.
[104] Galison, "The Ontology of the Enemy," 231 – 233.

(his) human-nonhuman status was blurred."[105] Canning takes this view of the enemy's human-machine status a step further by viewing the act of killing the operator as merely a consequence of destroying the machine.

Distance further facilitates killing. During the strategic bombardment campaign of World War II, being in an aircraft created physical distance between aircrew members and their victims. This distance enabled aircrews to deny "they were attempting to kill any specific individual."[106] In a similar manner, distance provided psychological protection for the civilian bombing victims who could deny they were the personal object of the attack.[107] Psychologically, the "*potential* of close-up, inescapable, *inter*personal hatred and aggression" is more terrifying than "the *presence* of inescapable, *im*personal death and destruction."[108]

Consider the firebombing of Dresden in July 1943. Regarding civilians killed in the attack, Grossman argues:

> If bomber crew members had had to turn a flamethrower on each one of these seventy thousand women and children or, worse yet, slit each of their throats, the awfulness and trauma inherent in the act would have been of such a magnitude that it simply would not have happened. But when it is done from thousands of feet in the air, where the screams cannot be heard and the burning bodies cannot be seen, it is easy.[109]

Distance provided the aircrew with the mental leverage to execute their macabre task. While intellectually these men understood "the horror of what they were doing," distance enabled them to deny it emotionally.[110]

Echoing Milgram's findings, Grossman argues distance lowers the innate resistance to killing.[111] Figure 11 provides a graphical

---

[105] Galison, "The Ontology of the Enemy," 233.
[106] Grosman, *On Killing*, 78.
[107] Grosman, *On Killing*, 78.
[108] Grosman, *On Killing*, 80.
[109] Grosman, *On Killing*, 100.
[110] Grosman, *On Killing*, 101 – 102.

representation of the relationship between distance and resistance to killing.

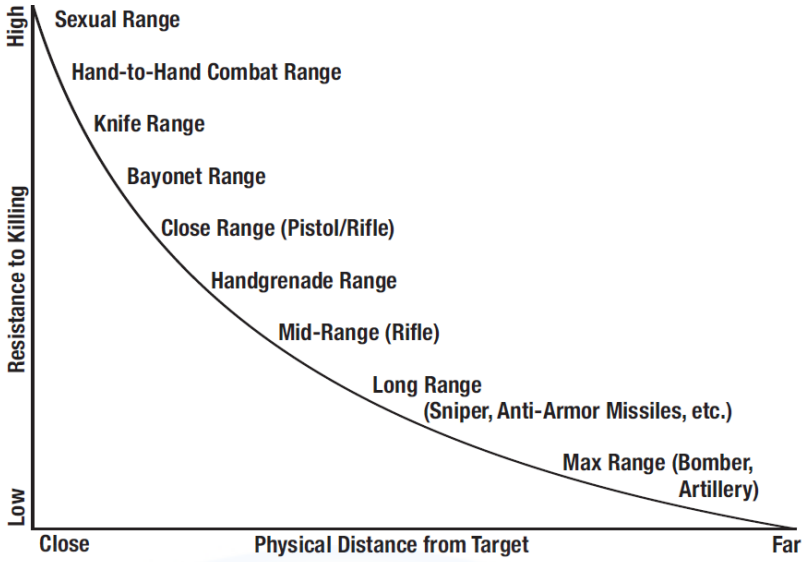

Figure 11: Resistance to Killing Based on Physical Distance

Source: Dave Grossman, On Killing, 98

Grossman defines maximum range as a "range at which the killer is unable to perceive his individual victims without using some form of mechanical assistance – binoculars, radar, periscope, or remote TV camera."[112] Maximum range would certainly describe the distance involved in killing with a lethal autonomous system.

The ability to kill from greater and greater distances has led to critiques of modern warfare categorizing it as postheroic, virtual war, and riskless war.[113] These characterizations hint at an underlying change in

---

[111] For an explanation of intimate violence, see Stathis N. Kalyvas, *The Logic of Violence in Civil War* (New York: Cambridge University, 2006), 330 – 362. Even with the intimate violence associated with civil wars, Kalyvas notes on page 350 a reluctance to act violently unless "someone else handles the gory details while shielding them."

[112] Grosman, *On Killing*, 107.

[113] For background information on postheroic warfare, see Edward Luttwak, *Strategy: The Logic of War and Peace* (Cambridge, MA: Harvard University, 2001), 68 – 74 and Edward N. Luttwak, "Toward Post-Heroic Warfare," *Foreign Affairs* 74, no. 3 (May/June 1995): 109 – 122. For virtual war, see Michael Ignatieff, *Virtual War: Kosovo and Beyond* (New York: Metropolitan Books, 2000). For riskless war, see Paul W. Kahn, "The Paradox of Riskless Warfare," *Yale Law School Faculty Scholarship Series* (2002), Paper 326.

how soldiers fight wars. No longer are physical courage and bravery the determinants of battle. Yet these critiques are not new. Improvements in artillery during the 1700s, such as those championed by the French artillerist Jean Baptiste Vacquette de Gribeauval, led to weapons that could "kill soldiers impersonally" at distances of more than half a mile, and thus created a capability that "offended deep-seated notions of how a fighting man ought to behave."[114] Artillery provided a capability that "seemed subversive of all that made a soldier's life heroic, admirable, worthy."[115] These early advancements in artillery technology merely foreshadowed the developments realized in modern warfare.

The notion of distance as it relates to the resistance to killing extends beyond physical distance and includes an emotional component. Cultural distance, moral distance, social distance, and mechanical distance create the emotional distance necessary to overcome the resistance to killing and are equally effective as physical distance in enabling the killer to deny the killing.[116] Cultural distance allows the dehumanization of the victim by the killer. The savage nature of fighting between the United States and Japan during World War II resulted in part from the cultural distance between the two belligerents.[117] General Sir Thomas Blamey, while visiting the battlefield at Buna, provided a paradigmatic example of this cultural distance when he asserted, "Fighting Japs is not like fighting normal human beings … The Jap is a little barbarian … We are not dealing with humans as we know them. We are dealing with something primitive. Our troops … regard them as vermin."[118] Wartime propaganda also created cultural distance. Even the propaganda cartoons of Theodor Geisel, better known as the beloved

---

[114] William H. McNeill, *The Pursuit of Power: Technology, Armed Force, and Society since A.D. 1000* (Chicago, IL: University of Chicago, 1982), 172.
[115] McNeill, *The Pursuit of Power*, 172.
[116] Grosman, *On Killing*, 158.
[117] Grosman, *On Killing*, 161 – 162.
[118] Quoted in John W. Dower, *War without Mercy: Race and Power in the Pacific War* (New York: Pantheon, 1986), 71.

children's author Dr. Seuss, created cultural distance through caricatures of the Japanese during World War II. Figure 12 below provides an example of Geisel's wartime work.[119]

Figure 12: World War II Propaganda Creating Cultural Distance

Source: Independent Lens, "The Political Dr. Seuss," http://www.pbs.org/independentlens/politicaldrseuss/seuss_fla.html (accessed 29 March 2012)

Moral distance emerges by condemning the enemy's guilt which deserves punishment and from the establishment of the legality and legitimacy of a cause. Battlecries such as "Remember the Alamo!" "Remember the *Maine*!" and "Remember Pearl Harbor!" reflect a form of moral distance.[120] Social distance establishes emotional distance between classes of people in a socially stratified environment.[121] In the military, social distance between officers and enlisted personnel creates a denial mechanism that allows officers to order their soldiers to their potential

[119] Cultural stereotypes created through propaganda were malleable. While wartime propaganda often portrayed the Japanese through aggressive looking simian caricatures, post-war depictions transformed the simian caricature into a "domesticated and even charming pet." See Dower, *War without Mercy*, 186.
[120] Grosman, *On Killing*, 164.
[121] Grosman, *On Killing*, 160.

death.[122] Mechanical distance stems from "the sterile Nintendo-game unreality of killing through a TV screen, a thermal sight, a sniper sight, or some other kind of mechanical buffer that permits the killer to deny the humanity of his victim."[123] In fact, the mechanical and physical distance created by killing with remotely piloted aircraft (RPA) led the United Nations to warn nations of the "risk of developing a 'Playstation' mentality to killing" and to encourage nations to ensure drone operators not exposed to battlefield risks respect and comply with international humanitarian law.[124]

Theoretically, the physical and mechanical distance created by a lethal autonomous system should reduce the resistance to killing, particularly when combined with a focus on destroying equipment rather than killing people. RPA provide an interesting case study for examining these effects. While humans operate RPA, these systems incorporate some autonomous elements and provide a technological gateway to a lethal autonomous system. RPA also provide insight into the effects of physical and mechanical distance on the resistance to killing.

Practical experience derived from studying RPA sensor operators paints a slightly more complex picture than the theory provides.[125] RPA sensor operators face the "highly unique challenge of providing continual support to combat operations in theaters of conflict while living and working in a peaceful environment and fulfilling domestic roles (i.e.,

---

[122] Grosman, *On Killing*, 169.
[123] Grosman, *On Killing*, 160.
[124] United Nations General Assembly, *Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, Philip Alston*, Addendum, *Study on Targeted Killings*, A/HRC/14/24/Add.6, 28 May 2010, 25 (para. 84).
[125] RPA sensor operators assist RPA pilots during missions. Key sensor operator duties include operation of aircraft sensors, assisting pilots in the development of weapons delivery tactics, and utilizing laser target marking systems to identify and mark targets for weapons delivery. For a complete list of RPA sensor operator duties, see Wayne Chappelle, Kent McDonald, and Raymond King, *Psychological Attributes Critical to the Performance of MQ-1 Predator and MQ-9 Reaper US Air Force Sensor Operators*, AFRL-SA-BR-TR-2010-0007 (Brooks City-Base, TX: Air Force Research Laboratory, 2010), 3 – 4.

spouse and parent) and responsibilities."[126] This challenge of balancing wartime and peacetime demands requires RPA sensor operators to compartmentalize their emotions daily and to have emotional stamina to balance these demands.[127] USAF researchers discovered an interesting dichotomy in performance based on whether operators focused on protecting US and allied forces or on destroying enemy combatants and assets. Those focusing on the former exhibited greater maturity and improved performance.[128]

Despite the physical and mechanical distance created by killing through an RPA, researchers found a small number of cases (approximately four or five) in which operators, after playing a role in the employment of weapons, voiced their discomfort with their roles and/or requested to leave the career field.[129] Finding a small group unwilling to participate in killing even with physical and mechanical distance is not surprising based on experimental results. In Milgram's experiments, 34 percent of the subjects were unwilling to administer the shock even in the remote case.[130] In this small subset of cases, the operators "performed their surveillance and reconnaissance duties well, but emotionally struggled with their role in taking the lives of others, regardless of the threat enemy combatants posed to US and allied forces."[131] The conflict these operators felt about their wartime duties did not manifest itself until the operators faced an actual combat situation

---

[126] Chappelle, *Psychological Attributes Critical to the Performance of MQ-1 Predator and MQ-9 Reaper US Air Force Sensor Operators*, 4.
[127] Chappelle, *Psychological Attributes Critical to the Performance of MQ-1 Predator and MQ-9 Reaper US Air Force Sensor Operators*, 19 – 20.
[128] Chappelle, *Psychological Attributes Critical to the Performance of MQ-1 Predator and MQ-9 Reaper US Air Force Sensor Operators*, 22.
[129] Chappelle, *Psychological Attributes Critical to the Performance of MQ-1 Predator and MQ-9 Reaper US Air Force Sensor Operators*, 22.
[130] Milgram, "Some Conditions of Obedience and Disobedience to Authority," 61 – 62.
[131] Chappelle, *Psychological Attributes Critical to the Performance of MQ-1 Predator and MQ-9 Reaper US Air Force Sensor Operators*, 22.

requiring the employment of weapons or until the operators became fully aware of the nature of their combat-related duties.[132]

One mechanism appearing to remind an operator with physical and mechanical distance of the gravity of these situations is an emotional tie to events. Human factors engineering provides the vehicle for these considerations during the design process. For an RPA, the radio can provide this emotional linkage between the operators and distant events. One RPA pilot reported, "When you're on the radio with a guy on the ground, and he is out of breath and you can hear the weapons fire in the background, you are every bit as engaged as if you were actually there."[133] This emotional connection could ground the operator in the situation enough to prevent the Playstation mentality while still providing enough distance to rationally evaluate the necessity to apply force.

Thus, while physical and mechanical distance can lower the resistance to killing, they do not always eliminate it. The design of systems that enable killing with significant physical and mechanical distance, particularly systems incorporating autonomy, also affects the willingness of humans to consent to the application of lethal force.

Automation Bias

Humans currently remain in the loop of automated decisions by retaining an override capability. In fact, Peter Singer of the Brookings Institute argues the military speaks of the human remaining in the loop so often that "it ends up sounding more like brainwashing than analysis."[134] Yet while human operators have the capability to override automatic weapons employment systems, they have demonstrated a

---

[132] Chappelle, *Psychological Attributes Critical to the Performance of MQ-1 Predator and MQ-9 Reaper US Air Force Sensor Operators*, 22.
[133] Major Matthew Morrison, RPA pilot, quoted in Christopher Drew, "Drones are Weapons of Choice in Fighting Qaeda," *The New York Times*, 17 March 2009, http://www.nytimes.com/2009/03/17/business/17uav.html?pagewanted=all (accessed 30 March 2012).
[134] Peter W. Singer, *Wired for War: The Robotics Revolution and Conflict in the 21st Century* (New York: Penguin, 2009), 124.

reluctance to do so. Thus, automation appears to increase mechanical distance and contribute to the denial defense mechanisms necessary to surmount the resistance to killing. Two tragic incidents illustrate this reluctance to override automatic systems: the 3 July 1988 downing of Iran Air Flight 655 and the 22 March 2003 downing of a Royal Air Force (RAF) Tornado GR4 by a Patriot missile.[135]

The *USS Vincennes*, a *Ticonderoga*-class Aegis cruiser, was operating in the Persian Gulf in July 1988. Iran Air Flight 655 was an Airbus 300 on a commercial flight from Bandar Abbas, Iran to Dubai, United Arab Emirates.[136] While Iran Air Flight 655 was flying a course typical for a commercial airliner, the Aegis system tagged the flight with a symbol that made it appear to be an Iranian F-14, leading the crew to categorize the aircraft as hostile rather than neutral.[137] Poor human-machine interface further contributed to the incident. While Iran Air Flight 655 was climbing away from the ship, the controllers perceived it was descending toward the ship. This misperception resulted because the display did not include the target aircraft's rate of altitude change, forcing controllers to calculate manually the altitude differential under combat situations.[138] Thus, despite radar data indicating that Iran Air Flight 655 was not an F-14, the crew of the *USS Vincennes* believed the Aegis system rather than interpreting the radar data for themselves. Rather than override the system's judgment, the crew authorized it to fire. This tragic unwillingness to override an automated system killed all 290 passengers on Iran Air Flight 655.[139]

---

[135] Peter W. Singer, "In the Loop? Armed Robots and the Future of War," *Defense Industry Daily*, 28 January 2009, http://www.defenseindustrydaily.com/In-the-Loop-Armed-Robots-and-the-Future-of-War-05267/ (accessed 22 March 2012).
[136] George C. Wilson, "Navy Missile Downs Iranian Jetliner," *Washington Post*, 4 July 1988, A1, http://www.washingtonpost.com/wp-srv/inatl/longterm/flight801/stories/july88crash.htm (accessed 22 March 2012).
[137] Singer, "In the Loop? Armed Robots and the Future of War."
[138] Mary L. Cummings, "Automation and Accountability in Decision Support System Interface Design," *The Journal of Technology Studies* 32 (Winter 2006), 23.
[139] Singer, "In the Loop? Armed Robots and the Future of War."

A similar unwillingness to override an automated system led to the downing of RAF Tornado GR4 ZG710 on 22 March 2003.[140] The aircraft was returning to Ali Al Salem Air Base in Kuwait from a mission in Iraq when a Patriot missile battery engaged and destroyed it. The Patriot missile system detected ZG710, but the system displayed the return as an anti-radiation missile heading directly toward the missile battery.[141] The Patriot system interrogated ZG710 for proper Identification Friend or Foe (IFF) codes but did not receive a response. Therefore, ZG710 met the necessary criteria for engagement by the Patriot battery in self-defense against an anti-radiation missile.[142]

The RAF Board of Inquiry specifically indicated the Patriot battery crew's trust of the system was a contributory factor to the incident. The Board stated, "Patriot crews are trained to react quickly, engage early and to trust the Patriot system. If the crew had delayed firing, ZG710 would probably have been reclassified as its flight path changed."[143] The US Defense Science Board Task Force assigned to evaluate the performance of the Patriot missile system during Operation IRAQI FREEDOM (OIF) concurred with the RAF assessment. The Task Force noted that trust in the system's automation would be necessary for an environment with many ballistic missile attacks–the situation for which engineers designed the system. During OIF, however, Patriot missile batteries engaged nine tactical ballistic missiles in an airspace congested

---

[140] Singer, "In the Loop? Armed Robots and the Future of War." The incident occurred at 2348 Zulu time on 22 March 2003, which corresponded to 0248 local time on 23 March 2003. Patriot missile batteries were also involved another fratricide incident in which they shot down a US Navy F/A-18 misidentified as a theater ballistic missile. Operator reluctance to override the system was also involved in this incident. See Department of Defense, *Patriot System Performance Report Summary*, Defense Science Board Task Force Report (Washington, DC: Department of Defense, January 2005), 2, http://www.acq.osd.mil/dsb/reports/ADA435837.pdf (accessed 3 May 2012).
[141] Ministry of Defense, *Aircraft Accident to Royal Air Force Tornado GR MK4A ZG710* (London, UK: Directorate of Air Staff, March 2004), 2.
[142] Ministry of Defense, *Aircraft Accident to Royal Air Force Tornado GR MK4A ZG710*, 2.
[143] Ministry of Defense, *Aircraft Accident to Royal Air Force Tornado GR MK4A ZG710*, 3.

with more than 41,000 coalition sorties.[144] Consequently, the automated Patriot system was operating outside its designed environment and did not perform as expected. Operators must understand that automation performs more effectively in some situations than others.[145] Caution should follow when operating an automated system out of its designed environment.

Despite the mistrust that critics have in allowing autonomous systems to make lethal decisions, the US military has already begun utilizing such systems. While these systems provide the human the capability of overriding the computer's judgment, these two incidents have shown how humans are reluctant to do so and illustrate the consequences of automation bias. Professor Mary Cumming from the Massachusetts Institute of Technology defines automation bias as the "tendency to disregard or not search for contradictory information in light of a computer generated solution that is accepted as correct."[146] Automation bias can lead to an operator "blissfully trusting the technology and abandoning responsibility for one's own actions."[147] It can also lead to "suboptimal decisions" with lethal consequences due to confusing or misleading recommendations.[148] While the intent of these systems is to reduce human errors and to improve decision-making

---

[144] Department of Defense, *Patriot System Performance Report Summary*, 2.
[145] John Hawley, "Practical Limits of Control: Lessons from the Patriot Vigilance Project," *Unmanned Platforms: Implication of Mission Autonomy for US Forces Conference* (Washington, DC: National Defense University, 19 May 2011), 10, http://www.ndu.edu/CTNSP/docUploaded/TFX_NDU%20Unmanned%20Platforms,%20Agenda,%20Bios,%20Presentations_May2011.pdf (accessed 3 May 2012).
[146] Cummings, "Automation and Accountability in Decision Support System Interface Design," 25. For more on the impact of automation bias in aviation, see Parasuraman, "Humans and Automation," 239 – 240.
[147] T.D. Sheridan, "Speculations on Future Relations Between Humans and Automation," *Automation and Human Performance,* ed. M. Mouloua (Mahwah, NJ: Lawrence Erlbaum Associates, 1996) quoted in Cummings, "Automation and Accountability," 25.
[148] Cummings, "Automation and Accountability," 24.

effectiveness, they can generate a perception by the operator that the automation is in charge.[149]

In fact, automation bias illustrates a Catch-22 for human operators resulting from the use of automation technology. Applying this technology creates the perception that the system can perform the task better than the human can. Yet the operator must remain in the loop to monitor the automated system's performance and override it when the system is wrong. Yet retaining an override capability carries an implicit assumption that the human operator can successfully identify situations requiring a veto over automated actions. Cognitive limitations and biases, however, make it difficult for human operators to meet this expectation.[150]

Blissful trust in technology has moral and ethical implications because it can diminish the operator's sense of responsibility and accountability.[151] A moral buffer can emerge that enables operators "to morally and ethically distance themselves from their actions."[152] When humans act through a "seemingly innocuous apparatus like a computer interface" and make lethal decisions with the click of a mouse, it creates a moral buffer enabling people to disassociate their responsibility from any resulting consequences.[153] A moral buffer can increase the ambiguity of responsibility for actions.[154] Furthermore, as an autonomous decision-making system continually makes accurate recommendations and as training reinforces the need to trust the system as in the Patriot battery example, operators come to rely on the system to make tough decisions. Rather than a recommended course of action, operators can begin to see solutions developed by the autonomous system as "a heuristic, a rule-of-thumb, which becomes the default

[149] Cummings, "Automation and Accountability," 25.
[150] Hawley, "Practical Limits of Control," 9.
[151] Cummings, "Automation and Accountability," 25.
[152] Cummings, "Automation and Accountability," 26.
[153] Cummings, "Automation and Accountability," 28 – 29.
[154] Cummings, "Automation and Accountability," 26.

condition, and hence a moral buffer."[155] As with any heuristic, overreliance can lead to errors.[156] Combining a moral buffer with physical and emotional distance has the potential to create a profound sense of detachment from the battlefield.[157] Therefore, as militaries develop these systems, officials must find a balance between encouraging trust in the system's recommended course of actions and independently evaluating such recommendations.

Engineers and system designers need to recognize that these moral and ethical issues emerge from human-machine interface decisions made during the development process.[158] Design decisions appearing to be simple engineering questions, such as what level of automation the system ought to have, can have unrecognized ethical and moral implications.[159] Even simple graphical user interface decisions can have unintentional ethical implications. Figure 13 illustrates how a design element in a computer interface could create a moral buffer.

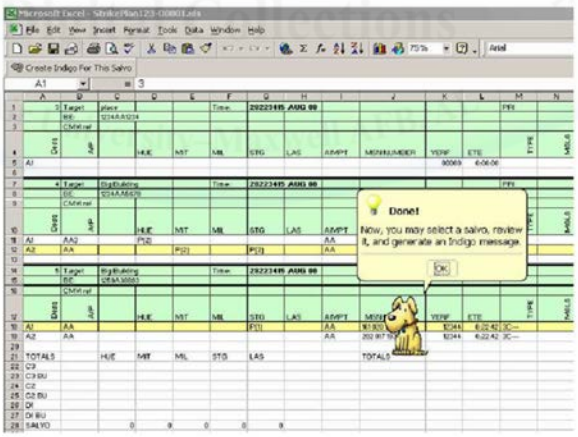

Figure 13: Excel Based Military Planning Tool

Source: Mary L. Cummings, "Automation and Accountability in Decision Support System Interface," 28.

[155] Cummings, "Automation and Accountability," 28.
[156] Parasuraman, "Humans and Automation," 239.
[157] Cummings, "Automation and Accountability," 27.
[158] Cummings, "Automation and Accountability," 23.
[159] Cummings, "Automation and Accountability," 24.

The figure captures a screenshot of a Microsoft Excel based planning tool for optimizing the flight path of a missile.[160] Including the "cheerful, almost funny graphic" of the dog announcing successful completion of mission planning could create a moral buffer by further detaching the planner's responsibility in the creation of lethal effects through an "innocuous medium."[161] While some might argue this interface is appropriate because it does not increase the stress facing the planner, trivializing the task to make it seem less distasteful is not the appropriate mechanism to reduce stress.[162]

Implementation of the responsibility advisor of Ronald Arkin's ethical architecture will face similar issues. Arkin avoids the use of superfluous graphics in a notional depiction of the responsibility advisor.[163] The responsibility advisor requires the operator to acknowledge granting final mission authorization for the autonomous system and to accept mission responsibility with a mouse click.[164] Once the mission is underway, the notional responsibility advisor displays the status of the system's permission to employ lethal force. In instances requiring the application of lethal force, the system provides the operator with a window to override the use of force.[165] Whether these efforts will be sufficient to avoid creating a moral buffer and to prevent operators from succumbing to automation bias, thus creating a Playstation mentality to killing, will largely depend on the design decisions made during system development. Human factors engineering will help inform these decisions. While engineers need to realize design decisions can have ethical and moral implications, they must also realize their creations likewise have moral and ethical consequences.

---

[160] Cummings, "Automation and Accountability," 28.
[161] Cummings, "Automation and Accountability," 28.
[162] Cummings, "Automation and Accountability," 28.
[163] Arkin, *Governing Lethal Behavior in Autonomous Robots*, 196 – 209.
[164] Arkin, *Governing Lethal Behavior in Autonomous Robots*, 201.
[165] Arkin, *Governing Lethal Behavior in Autonomous Robots*, 204 – 206.

Assessing the Implications of Technological Development

Technological advancements create not only "marvelous new machines and new methods of controlling nature" but also "new problems in social and political development and set in motion conflicting forces which must eventually be resolved."[166] Humanity has struggled with this problem throughout history. Reflecting upon his technical work during World War II, MIT mathematics professor Norbert Wiener lamented he stood in a "moral positions which is, to say the least, not very comfortable."[167] Recognizing humanity could use technological developments for good or for evil, Wiener conceded scientists and engineers could only hand over their creations "into the world that exists about us, and this is the world of Belsen and Hiroshima. We do not even have the choice of suppressing these new technological developments."[168] Developing new technology creates an obligation to examine the moral and ethical implications of these creations. Unfortunately, humanity often appears reluctant to grapple with these issues. Speaking about the immense destructive power of nuclear weapons, General Omar Bradley deplored the ability of technological advancements to outpace moral and ethical thought.

> With the monstrous weapons man already has, humanity is in danger of being trapped in this world by its moral adolescence. Our knowledge of science has clearly outstripped our capacity to control it…Man is stumbling blindly through a spiritual darkness while toying with the precarious secrets of life and death. The world has achieved brilliance without wisdom, power without conscience. Ours is a world of nuclear giants and ethical infants. We know

---

[166] Watson O'D. Pierce, *Air War: Its Psychological, Technical, and Social Implications* (New York: Modern Age Books, 1939), 215.
[167] Norbert Wiener, *Cybernetics, or Control and Communication in the Animal and the Machine*, 2d ed. (New York: MIT Press, 1961), 28.
[168] Wiener, *Cybernetics*, 28. Belsen refers to the Nazi concentration camp Bergen-Belsen.

> more about war than we know about peace, more about killing than we know about living.[169]

Since society typically views these moral and ethical dilemmas as social and political questions, a democratic society must debate these issues. As the Manhattan Project ended, the Smyth Report emphasized the need for the public to discuss the social and political issues raised with nuclear weapons.

> The future possibilities of such explosives are appalling, and their effects on future wars and international affairs are of fundamental importance. Here is a new tool for mankind, a tool of unimaginable destructive power. Its development raises many questions that must be answered in the near future.
>
> Because of the restrictions of military security there has been no chance for Congress or the people to debate such questions. They have been seriously considered by all concerned and vigorously debated among the scientists, and the conclusions reached have been passed along to the highest authorities. These questions are not technical questions; they are political and social questions, and the answers given to them may affect all mankind for generations…In a free country like ours, such questions should be debated by the people and decisions must be made by the people through their representatives…The people of the country must be informed if they are to discharge their responsibilities wisely.[170]

Wiener echoed the sentiments of the Smyth Report when he concluded "the best we [scientists] can do is to see that a large public understands the trend and the bearing of the present work."[171]

Peter Singer has decried the difficulty in discussing the ethical and moral ramifications of emerging technology in war.[172] Singer argues one

---

[169] General Omar N. Bradley, "An Armistice Day Address" (address, Chamber of Commerce, Boston, MA, 10 November 1948). The text is available at http://www.opinionbug.com/2109/armistice-day-1948-address-general-omar-n-bradley (accessed 21 March 2012).
[170] Henry D. Smyth, *Atomic Energy for Military Purposes* (York, PA: Maple Press, 1945), 226. Accessed at http://www.archive.org/details/atomicenergyform00smytrich (8 February 2012).
[171] Wiener, *Cybernetics*, 28.

of the complicating factors is the disconnect among disparate academic fields. Going from one discipline into another can be "like crossing into a foreign land."[173] People become uncomfortable when reaching beyond their area of expertise, as it exposes them to unfamiliar territory and language. Capturing this sentiment, one robotics professor Singer interviewed stated, "Having discussions about ethics is very difficult because it requires me to put on a philosopher's hat, which I don't have."[174] Therefore, people remain in their own lanes. Moral and ethical responsibility diffuses further. Researchers create a moral buffer by telling themselves humanity can utilize their developments for good or evil so responsibility for actions outside the laboratory lies beyond the researchers.[175]

Considering the moral and ethical implications after contributing to technological advancements is too late. Norbert Wiener illustrates the consequences of waiting too long to ask these questions. In a letter to the philosopher Giorgio de Santillana, a despondent Wiener bemoaned, "Ever since the atomic bomb fell I have been recovering from an acute attack of conscience as one of the scientists who has been doing war work and who has seen his war work a[s] part of a larger body which is being used in a way of which I do not approve and over which I have absolutely no control."[176]

Those developing lethal autonomous systems are at risk of donning ethical blinders while remaining in their lane. Sixty percent of the top twenty-five stakeholders in the robotics field simply answered "No" when asked whether they foresaw any social, ethical, or moral problems arising

---

[172] Peter W. Singer, "The Ethics of Killer Applications: Why is it so Hard to Talk About Morality When it Comes to New Military Technology," *Journal of Military Ethics* 9, no 4 (December 2010): 299 – 312.
[173] Singer, "The Ethics of Killer Applications," 301.
[174] Singer, "The Ethics of Killer Applications," 301.
[175] Singer, *Wired for War*, 175.
[176] Norbert Wiener quoted in Peter Galison, "The Ontology of the Enemy: Norbert Wiener and the Cybernetic Vision," *Critical Inquiry* 21, no. 1 (Autumn 1994), 253.

from the continued development of unmanned systems.[177] Yet, it is possible to foster these necessary discussions. The Human Genome Project, for example, earmarks five percent of its annual budget to foster discussions on the project's ethical, legal, and social implications.[178]

Confronting war's harsh realities as opposed to an idealized notion of war illuminates the "double-edged sword of technology."[179] Superior military technology provides a significant advantage in warfare.[180] Consequently, nations pursue development of the "proverbial technological silver bullet."[181] In fact, Colin Gray, the renowned British strategic thinker and Professor of International Relations and Strategic Studies at the University of Reading, identifies the engineering style and the technical fix as a characteristic of the American strategic culture. This characteristic gives rise to the notion that "American know-how will find a solution–generally technical–to every problem."[182] These choices have costs and ethical implications, and unfortunately, "silver bullet-technological solutions for ethics" do not exist.[183]

Mao Tse-Tung chided Duke Hsiang of Sung for his "asinine ethics."[184] The Duke allowed the Ch'u troops to complete a river crossing before ordering his army to attack despite pleadings from his ministers to commence the attack while the enemy troops were conducting their river crossing.[185] That ethics hinder the conduct of warfare presents a powerful, popular, and persistent narrative. Yet the historic record

---

[177] Singer, "The Ethics of Killer Applications," 301.
[178] Singer, "The Ethics of Killer Applications," 302.
[179] Singer, "The Ethics of Killer Applications," 303.
[180] Kenneth P. Werrell, *Chasing the Silver Bullet: U.S. Air Force Weapons Development from Vietnam to Desert Storm* (Washington, DC: Smithsonian Books, 2003), 2.
[181] Werrell, *Chasing the Silver Bullet*, 8.
[182] Colin S. Gray, *Explorations in Strategy* (Westport, CT: Praeger, 1996), 90.
[183] Peter W. Singer, "The Ethics of Killer Applications: Why is it so Hard to Talk About Morality When it Comes to New Military Technology," *Journal of Military Ethics* 9, no 4 (December 2010), 304.
[184] Mao Tse-Tung, "On Protracted War," *Selected Works of Mao Tse-Tung*, vol. 2, (Peking, China: Foreign Language Press, 1965; new imprint Digital Reprints 2007), 166.
[185] Sun Tzu, *The Illustrated Art of War*, trans. Samuel Griffith (New York: Oxford University, 2005), 181 – 182. The Duke's forces lost the ensuing battle.

illustrates that professional armies "almost always triumph over those who are willing to behave like barbarians."[186] Colin Gray argues in the three great wars of the Twentieth Century–World War I, World War II, and the Cold War–the victors "enjoyed the advantages of an ethically compelling story."[187] Thus, ethical behavior becomes a component of achieving military victory. Ethical behavior in warfare requires all involved, from those developing new technologies to soldiers, to assess the moral and ethical implications of new battlefield technologies and new methods of fighting.

Conclusion

The Scientific Advisory Board Report *Toward New Horizons: Science, Key to Air Supremacy* championed by Dr. Theodore von Karman and General Hap Arnold helped establish the central role of advanced technology in the US Air Force. Even at the end of World War II, von Karman's team foresaw the coming importance of pilotless aircraft, a vision now reaching fruition. Along the way, the conceptualization of a human-machine symbiosis has promised the ability to augment strengths and diminish weaknesses in human decision-making. Yet, achieving this symbiosis has proven difficult.

The design of the human-machine interface influences the nature of interactions. Human factors engineering provides insights into how best to divide tasks between humans and machines. The resulting interactions have ethical and moral implications that may unintentionally go unnoticed.

Consenting to the execution of lethal force through a machine creates mechanical distance and lowers the human resistance to killing. Additionally, relying on automated decision-making mechanisms can lead to automation bias, a moral buffer resulting in humans deferring

[186] Singer, "The Ethics of Killer Applications," 309.
[187] Colin S. Gray, *Modern Strategy* (New York: Oxford University, 1999), 73.

hard choices to machine judgment. Interface design can also create a moral buffer, further enabling operators to distance themselves from the moral and ethical consequences of their actions.

Discussing the moral and ethical consequences of these technological advancements is difficult. This difficulty emerges in part due to the necessity of crossing out of narrowly constructed lanes of expertise. While we are not all ethicists, we must exercise judgment and put on our philosopher's hat to consider the implications of technological advancements and our involvement in these advances. The *USAF UAS Flight Plan* asserts the decision to allow a machine to make lethal decision in combat is "contingent upon political and military leaders resolving legal and ethical questions."[188] All involved in the process have a duty to inform the discussion on these questions.

Examining the legal and ethical issues surrounding lethal autonomous systems illuminates not only the technical obstacles engineers must overcome but also the repercussions of choices made during the design process. Informed design decisions help prevent the creation of systems enabling a Playstation mentality to killing. Yet the fundamental question remains: why employ an autonomous capability? By answering this basic question, the decision to use autonomy becomes a conscious choice, forcing the consideration of advantages and disadvantages, rather than the next technological development whose implications are considered only after the fact.

[188] Department of the Air Force, *United States Air Force Unmanned Aircraft Systems Flight Plan 2009 – 2047* (Washington, DC: Headquarters, United States Air Force, 18 May 2009), 41.

Conclusion

One of the lamentable principles of human productivity is that it is easier to destroy than to create.

—Thomas C. Schelling
Arms and Influence

Giulio Douhet, the early Italian airpower theorist, recognized that aerial attacks against an enemy's civilian population would cause a lamentable loss of life. He assuaged concerns about terror bombing by arguing these attacks might hasten the war's end and "may in the long run shed less blood."[1] As American airpower theorists developed strategic bombing doctrine between the World Wars at the Air Corps Tactical School (ACTS), they were aware of the potential necessity for attacking political capitols and centers of population. These officers believed, however, that "political considerations will govern, and warfare of terrorization will probably be conducted only as a matter of reprisal."[2] While ethical and legal questions remained about aerial bombardment, ACTS thinking evolved into the industrial fabric theory, paving a path forward to undermine the adversary's morale while side-stepping the associated ethical questions.[3]

Those working on autonomous systems today find themselves in an analogous situation. Political and military leaders have not resolved legal and ethical questions about these systems. International law is silent on the legality of autonomous systems. Questions about the legality of these systems must turn to the underlying principles of the law of armed conflict: necessity, proportionality, discrimination, and

---

[1] Giulio Douhet, *The Command of the Air*, ed. Joseph Harahan and Richard Kohn (Tuscaloosa, AL: University of Alabama, 2009), 61.
[2] ACTS, "International Aerial Regulations" text, 1933-34, in AFHRC, decimal file no. 248.101-16, 1933-34, 31.
[3] Tami Biddle, *Rhetoric and Reality: The Evolution of British and American Ideas About Strategic Bombing, 1914 – 1945*, (Princeton, NJ: Princeton University, 2002), 160.

humaneness.[4] In general, lethal autonomous systems do not appear to violate these principles, yet their ultimate legality will depend upon the details of implementation. Some argue against these systems on moral and ethical grounds based on a perceived inability to apply the principle of discrimination and lack of accountability for the actions of these systems.[5] Others assert autonomous systems will make it easier for political leaders to lead the nation into war, that the technology will have unintended consequences, or that these systems will adversely affect the military ethos.[6] Compelling arguments, however, also exist for the use of these systems. To aid policy-makers in resolving the questions surrounding lethal autonomous systems, the military must understand the arguments both for and against these systems.

While these ethical and legal questions persist and remain largely unanswered, work is continuing on the development of lethal autonomous systems, just as work progressed on early strategic bombardment aircraft despite nagging and unresolved concerns about its ethical and legal implications. The science fiction of Isaac Asimov's Three Laws of Robotics provoked thought on how to develop ethical reasoning for a machine, yet the Three Laws do not provide a viable solution.[7] Three general approaches for creating ethical reasoning in autonomous systems exist: a top-down approach that encodes a particular ethical

---

[4] Article 36, Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts (Protocol I), 8 June 1977, 1125 U.N.T.S. 25.

[5] For examples of these arguments, see Noel Sharkey, "Grounds for Discrimination: Autonomous Robot Weapons," *RUSI Defense Systems – Challenges of Autonomous Weapons* (October 2008): 86 - 89, and Robert Sparrow, "Killer Robots," *Journal of Applied Philosophy* 24, no. 1 (February 2007): 62 – 77.

[6] For examples of these arguments, see Peter Asaro, "How Just Could a Robot War Be?" in *Current Issues in Computing and Philosophy*, ed. Adam Briggle, Katinka Waelbers, and Philip Brey, (Fairfax, VA: IOS Press, 2008): 50 - 64, Maj Daniel L. Davis, "Who Decides: Man or Machine?" *Armed Forces Journal*, November 2007, and Lt Col Michael Contratto, "The Decline of the Military Ethos and Profession of Arms: An Argument Against Autonomous Lethal Engagements," (Air War College paper, Air University, 2011).

[7] Ronald C. Arkin, *Governing Lethal Behavior in Autonomous Robots* (New York: CRC Press, 2009), 48.

theory, a bottom-up approach that enables a system to learn morality based on experience, and a hybrid approach that combines the top-down and bottom-up approaches.[8] The top-down approach gives rise to concerns of slave morality, while the bottom-up approach gives rise to systems running amok by learning to override restraints.

Ronald Arkin from the Georgia Institute of Technology has proposed a specific ethical architecture for an autonomous lethal system with four components–an ethical governor, an ethical behavior control, an ethical adaptor, and a responsibility advisor.[9] His approach addresses concerns about accountability but relies on further technical advances to address discrimination and proportionality concerns. John Canning of the Naval Surface Warfare Division has proposed an approach that would have autonomous systems target things rather than people. While this approach involves legal contortions, it possesses historical precedent and makes discrimination somewhat easier. Regardless of the approach chosen, trust that the autonomous system will operate as designed is essential. Lack of a means to accomplish verification and validation of these systems presents a significant barrier.[10] Scientists, however, have begun developing the necessary techniques and procedures to overcome this barrier.

Technological advances have blurred the boundary between human and machine.[11] Recognizing that humans excel at certain tasks and machines at others highlights the importance of developing an

---

[8] Patrick Lin, George Bekey, and Keith Abney, *Autonomous Military Robotics: Risk, Ethics, and Design*, Office of Naval Research Report N00014-07-1-1152 (San Luis Obispo, CA: California Polytechnic State University, 20 December 2008), 27.
[9] Arkin, *Governing Lethal Behavior in Autonomous Robots*, 125 – 126.
[10] Werner J.A. Dahm, *Technology Horizons: A Vision for Air Force Science and Technology During 2010 – 2030*, AF/ST-TR-10-01-PR, (Washington, DC: Department of the Air Force, 15 May 2010), 42.
[11] David A. Mindell, *Between Human and Machine: Feedback, Control, and Computing before Cybernetics* (Baltimore, MD: Johns Hopkins University, 2002), 2.

appropriate division of tasks.[12] In fact, system design influences the nature of human-machine interactions and can have moral implications. Consequently, human factors engineering helps assess the proper assignment of tasks. With the blurring of the line between human and machine, humans have begun ceding some authority for lethal-decision making to machines. While human operators remain in the loop for weapons release decisions, they have a reluctance to override the system's decisions due to automation bias.[13] Furthermore, the system interface design has the potential to create a moral buffer for the operator by detaching the operator from the consequences of actions.[14] Due to the potential effects of lethal autonomous systems, those involved in developing these systems must don their philosopher's hat and probe the difficult moral and ethical implications of their work.[15]

The road to a fully autonomous system capable of employing lethal force remains long. Examining the potential legal and moral implications of these systems informs their development. It highlights technical roadblocks and the implications of design choices. The near-term path forward should focus on developing systems more capable of advising the human operator on how best to employ lethal force. Yet the design of this process must remain centered on the operator to prevent reliance on automated processes and the abdication of moral responsibility. Training must emphasize where the automation and autonomy are robust and where they are not to aid operators in understanding when overriding the system might be necessary. An operator-centered focus will continue the

---

[12] Paul M. Fitts, *Human Engineering for an Effective Air-Navigation and Traffic-Control System* (Washington, DC: National Research Council, March 1951), 6 – 8.
[13] Mary L. Cummings, "Automation and Accountability in Decision Support System Interface Design," *The Journal of Technology Studies* 32 (Winter 2006), 25.
[14] Cummings, "Automation and Accountability," 28.
[15] Peter W. Singer, "The Ethics of Killer Applications: Why is it so Hard to Talk About Morality When it Comes to New Military Technology," *Journal of Military Ethics* 9, no 4 (December 2010), 301.

blurring of the line between human and machine, taking advantage of the relative strengths of both and leading to a more effective symbiosis.

The eminent historian Lynn White, Jr., argues "a new device merely opens a door; it does not compel one to enter. The acceptance or rejection of an invention, or the extent to which its implications are realized if it is accepted, depends quite as much upon the condition of society, and upon the imagination of its leaders, as upon the nature of the technological item itself."[16] Given the American culture, the nation appears open to the development of a lethal autonomous system. A central tenet of the modern American way of war has been the "reliance on advanced technology."[17] Indeed, the American military subscribes to the belief that "superior weapons give their users an advantage favoring victory."[18] Another trend in American warfare has been the use of humanity as a weapon of war, whereby the nation derives a strategic advantage from demonstrating the extent to which it goes to minimize suffering to non-combatants during armed conflict.[19] Thus, the development of an autonomous system capable of employing lethal force has the potential to solve a beguiling conundrum facing American policy-makers. These systems may present a tool to take military action for policy aims short of vital threats to American national interests without risking American lives while minimizing non-combatant casualties.

Yet political and military leaders must understand the implications of utilizing these systems. As former Deputy Assistant Secretary of Defense for Policy Planning Thomas Mahnken asserts, "America's traditional reliance upon technology in war is certainly no recipe for

---

[16] Lynn White, Jr., *Medieval Technology and Social Change* (New York: Oxford, 1964), 28.
[17] Thomas G. Mahnken, Technology and the American Way of War Since 1945, (New York: Columbia University, 2008), 5.
[18] Irving B. Holley, Jr., *Ideas and Weapons* (Washington, DC: Government Printing Office, 1997), 175.
[19] Reuben E. Brigety II, *Ethics, Technology, and the American Way of War: Cruise Missiles and US Security Policy* (New York: Routledge, 2007), 35 – 40.

success. Technology is a poor substitute for strategic thinking."[20] Martin van Creveld also highlights this point by arguing, "Any given technology has very strict limits. Often the critical factor is less the type of hardware available than the way it is put to use."[21] The development of autonomous lethal systems will necessitate new strategic analysis on how best to employ these systems. Like any other technological system, they will have distinct advantages and disadvantages. Placing too much faith in advanced technology can blind one to the potential countermeasures of an opponent.[22] A lethal autonomous system could enable the nation to "see war as a surgical scalpel and not a bloodstained sword."[23]

Langdon Winner, a political theorist at Rensselaer Polytechnic who focuses on the social and political issues accompanying modern technological change, cautions, "Although virtually limitless in their power, our technologies are tools without handles."[24] Furthermore, technological artifacts have political qualities. The intricacies of a system's design yield specific, logical consequences that extend beyond the system's professed intent.[25] For example, the designer of roads and bridges in New York specifically designed low-hanging overpasses on Long Island in order to restrict access to the parkways to the upper and middle classes who owned cars. Since the low-hanging overpasses restricted the ability of buses to use the parkway, the design effectively prevented low-income groups who relied on public transportation from reaching Jones Beach–an acclaimed public park only accessible via the

---

[20] Mahnken, *Technology and the American Way of War Since 1945*, 6.
[21] Martin Van Creveld, *Command in War (*Cambridge, MA: Harvard University, 1985), 231.
[22] Joseph Nye, *The Future of Power* (New York: Public Affairs, 2011), 36.
[23] Michael Ignatieff, *Virtual War: Kosovo and Beyond* (New York: Metropolitan Books, 2000), 215.
[24] Langdon Winner, *Autonomous Technology Technics-out-of-Control as a Theme in Political Thought* (Boston, MA: MIT, 1978), 29.
[25] Langdon Winner, "Do Artifacts Have Politics?" *Daedalus* 109, no. 1 (Winter 1980), 134.

parkway.[26] Of course, not all consequences are so nefarious. Nevertheless, the design of a technological system provides a mechanism to establish "patterns of power and authority."[27] For example, the lethality of nuclear weapons necessitated the establishment of a centralized and rigid hierarchical chain of command to make this control system predictable. In short, these weapons required an authoritarian management system. Consequently, preventing the spillover of this authoritarian nature of the management system for these weapons into society as a whole became a particular challenge for democratic societies possessing nuclear weapons.[28] Given the power and the range of flexibility inherent in the design of a technological system, system developers must anticipate and understand the consequences of the designs and arrangements they choose.[29] Those involved in the development of lethal autonomous systems must keep these cautions in mind.

[26] Winner, "Do Artifacts Have Politics?" 123 – 124.
[27] Winner, "Do Artifacts Have Politics?" 134.
[28] Winner, "Do Artifacts Have Politics?" 131.
[29] Winner, "Do Artifacts Have Politics?" 134.

Glossary

ACL	Autonomous Control Levels
ACTS	Air Corps Tactical School
AFHRC	Air Force Historical Research Center
AFRL	Air Force Research Laboratory
AI	Artificial Intelligence
ALFUS	Autonomy Level for Unmanned Systems
AOC	Air Operations Center
AP	Additional Protocol
AU	Air University
AWPD	Air War Plans Division
CAS	Close Air Support
CENTCOM	US Central Command
CRAM	Counter Rocket Artillery Mortar
DARPA	Defense Advanced Research Projects Agency
DOD	Department of Defense
DTC	Defense Technology Centre
HUD	Head-Up Display
ICRAC	International Committee for Robot Arms Control
IEEE	Institute of Electrical and Electronics Engineers
IFF	Identification Friend or Foe
IHL	International Humanitarian Law
ISR	Intelligence, Surveillance, and Reconnaissance

MIT	Massachusetts Institute of Technology
MOD	Ministry of Defense
NASA	National Aeronautics and Space Administration
NATO	North Atlantic Treaty Organization
NIST	National Institute of Standards and Technology
OIF	Operation IRAQI FREEDOM
OODA	Observe, Orient, Decide, and Act
OUSD (AT&L)	Office of the Under Secretary of Defense (Acquisition, Technology and Logistics)
RAF	Royal Air Force
RPA	Remotely Piloted Aircraft
SADL	Situational Awareness Data Link
SANDF	South African National Defense Force
SEAS	Systems Engineering for Autonomous Systems
UAS	Unmanned Aircraft System
UN	United Nations
UMS	Unmanned System
USAF	United States Air Force
V&V	Verification and Validation

Bibliography

Academic Papers

Alexander, Larry and Michael Moore. "Deontological Ethics." In *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*, edited by Edward N. Zalta. http://plato.stanford.edu/archives/fall2008/entries/ethics-deontological (accessed 20 February 2012).

Anderson, Kenneth. "Efficiency in Bello and ad Bellum: Targeted Killing Through Drone Warfare." *American University Working Paper Series*, 23 September 2011.

Arkin, Ronald C. "Robots in Warfare." *IEEE Technology and Society Magazine*, Spring 2009: 30 – 33.

Arkin, Ronald C., and Patrick Ulam. "An Ethical Adaptor: Behavioral Modification Derived from Moral Emotions." *2009 IEEE International Symposium on Computational Intelligence in Robotics and Automation*, December 2009.

Arrow, Kenneth J. "Uncertainty and the Welfare Economics of Medical Care." *The American Economic Review* 53, no. 5 (Dec 1963): 941 – 973.

Asaro, Peter. "How Just Could a Robot War Be?" In *Current Issues in Computing and Philosophy*, edited by Adam Briggle, Katinka Waelbers, and Philip Brey, 50 – 64. Fairfax, VA: IOS Press, 2008.

Asaro, Peter. "What Should We Want From a Robot Ethic?" *International Review of Information Ethics* 6 (December 2006): 9 – 15.

Banks, Sheila B. and Carl S. Lizza. "Pilot's Associate: A Cooperative, Knowledge-Based System Application." *IEEE Intelligent Systems and their Applications* (June 1991): 18 – 29.

Belkin, Brenda L. and Robert F. Stengel. "Cooperative Rule-Bases Systems for Aircraft Control." Paper presented. *Proceedings of the 26th Conference on Decision and Control.* Los Angeles, CA, December 1987.

Bostrom, Nick. "Ethical Issues in Advanced Artificial Intelligence." In *Science Fiction and Philosophy: From Time Travel to Superintelligence*, edited by Susan Schneider, 277 - 284. Oxford, UK: Blackwell Publishing, 2009.

Brat, Guillaume and Ari Jonsson. "Challenges in Verification and Validation of Autonomous Systems for Space Exploration." Paper presented. *2005 IEEE International Joint Conference on Neural Networks*, Montreal, Quebec, Canada, 31 July – 4 August 2005.

Brunstetter, Daniel and Megan Braun. "The Implications of Drones on the Just War Tradition." *Ethics & International Affairs* 25, no. 3 (Fall 2011), 348 – 352

Builder, Carl. "Service Identities and Behavior." In *American Defense Policy*, edited by Peter L. Hays, Brenda J. Vallance, and Alan R. Van Tassel, 108 – 122. Baltimore, MD: Johns Hopkins University, 1997.

Canning, John S. "A Concept of Operations for Armed Autonomous Systems," Paper presented. *National Defense Industrial Association 3rd Annual Disruptive Technology Conference "Seeking the Capability Before the Capability is the Surprise,"* Washington, DC, 6 – 7 September 2006.

Canning, John S. "Weaponized Unmanned Systems: A Transformational Warfighting Opportunity, Government Roles in Making it Happen." Paper Presented. *American Society of Naval Engineers Engineering the Total Ship Symposium*, Falls Church, VA, 23 – 25 September 2008.

Clarke, Roger. "Asimov's Laws of Robotics: Implications for Information Technology – Part II." *IEEE Computer* 27, no. 1 (Jan 1994), 57 – 66.

Clough, Bruce. "Metrics, Schmetrics! How the Heck do you Determine a UAV's Autonomy Anyway?" Paper presented. *2002 Performance Metrics for Intelligent Systems Workshop*, Gaithersburg, MD, 13 – 15 August 2002.

Cohen, Eliot. "The Mystique of US Air Power." *Foreign Affairs* 73, no. 1 (Jan – Feb 1994): 109 – 124.

Contratto, Lt Col Michael. "The Decline of the Military Ethos and Profession of Arms: An Argument Against Autonomous Lethal Engagements." Master's thesis, Air War College, 2011.

Cummings, Mary L. "Automation and Accountability in Decision Support System Interface Design," *The Journal of Technology Studies* 32 (Winter 2006): 23 – 31.

Cummings, Mary L., Andrew Clare, and Christin Hart. "The Role of Human-Automation Consensus in Multiple Unmanned Vehicle Scheduling," *Human Factors: The Journal of the Human Factors and Ergonomics Society* 52, no. 1 (February 2010): 17 – 27.

Cummings, Mary L. and K.M. Thornburg, "Paying Attention to the Man Behind the Curtain," *IEEE Pervasive Computing* 10 (Jan – Mar 2011), 58 – 62. http://www.web.mit.edu/aeroastro/labs/halab/papers/Final_Curtain.pdf (accessed 21 March 2012).

Dancy, Jonathan. "Moral Particularism." In *The Stanford Encyclopedia of Philosophy (Spring 2009 Edition)*, edited by Edward N. Zalta. http://plato.stanford.edu/archives/spr2009/entries/moral-particularism (accessed 20 February 2012).

Dembe, Allard E. and Leslie I. Boden. "Moral Hazard: A Question of Morality?" *New Solutions: A Journal of Environmental and Occupational Health Policy* 10, no. 3 (2000): 257 – 279.

Galison, Peter. "The Ontology of the Enemy: Norbert Wiener and the Cybernetic Vision." *Critical Inquiry* 21, no. 1 (Autumn 1994): 228 – 266.

Gillespie, Tony and Robin West. "Requirements for Autonomous Unmanned Air Systems Set by Legal Issues." *The International C2 Journal* 4, no. 2 (2010): 1 – 32.

Gulpinar, Nalan and Ethem Canakoglu. “Robust Team Coordination and Decision Making under Uncertainty.” Paper presented. *5th Systems Engineering for Autonomous Systems Defense Technology Centre Technical Conference*, Edinburgh, UK, 14 – 15 July 2010. Paper B6.

Gupta, Pramod and Johann Schumann. “A Tool for Verification and Validation of Neural Network Based Adaptive Controllers for High Assurance Systems.” Paper presented. *Eighth IEEE International Symposium on High Assurance Systems Engineering*, Tampa, FL, 25 – 26 March 2004.

John. Hawley. “Practical Limits of Control: Lessons from the Patriot Vigilance Project.” Paper presented. *Unmanned Platforms: Implication of Mission Autonomy for US Forces Conference*, National Defense University, Washington, DC, 19 May 2011. http://www.ndu.edu/CTNSP/docUploaded/TFX_NDU%20Unmanned%20Platforms,%20Agenda,%20Bios,%20Presentations_May2011.pdf (accessed 3 May 2012).

Hopper, Major Aaron M. “The Future of Autonomy in U.S. Air Force Unmanned Air Systems: Toward a Strategy for Growth.” Master’s thesis, Air Command and Staff College, 2011.

Hughes, Thomas P. “The Evolution of Large Technological Systems.” In *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*, edited by Wiebe E. Bijker, Thomas P. Hughes, and Trevor Pinch. Cambridge, MA: MIT Press, 1989.

Hursthouse, Rosalind. “Virtue Ethics.” *The Stanford Encyclopedia of Philosophy (Winter 2010 Edition)*, edited by Edward N. Zalta. http://plato.stanford.edu/archives/win2010/entries/ethics-virtue (accessed 20 February 2012).

Jenks, Chris, Lt Col. “Law from Above: Unmanned Aerial Systems, Use of Force, and the Law of Armed Conflict.” *North Dakota Law Review* 85, no. 3 (2009): 649 – 671.

Jones, Randolph M., John E. Laird, Paul E. Nielsen, Karen J. Coulter, Patrick Kenny, and Frank V. Koss. “Automated Intelligent Pilots for Combat Flight Simulation.” *AI Magazine* 20, no. 1 (Spring 1999): 27 – 41.

Kahn, Paul W. “The Paradox of Riskless Warfare.” *Yale Law School Faculty Scholarship Series* (2002). Paper 326.

Larkin, Maj Sean P. “Air-to-Ground Fratricide Reduction Technology: An Analysis.” Master’s thesis, Marine Corps Command and Staff College, 2005.

Lee, Caitlin H. “Embracing Autonomy: The Key to Developing a New Generation of Remotely Piloted Aircraft for Operations in Contested Air Environments.” *Air and Space Power Journal* 24, no. 4 (Winter 2011): 76 – 88.

Licklider, J.C.R. “Man-Computer Symbiosis.” *IRE Transactions on Human Factors in Electronics* 1 (March 1960): 4 – 11.
Lin, Patrick, George Bekey, and Keith Abney. “Robots in War: Issues of Risk and Ethics.” In *Ethics and Robotics*, edited by Rafael Capurro and Michael Nagenborg, 49 - 67. Heidelberg: AKA Verlag, 2009.
Luttwak, Edward N. “Toward Post-Heroic Warfare.” *Foreign Affairs* 74, no. 3 (May/June 1995): 109 – 122
Marx, Leo. “The Idea of ‘Technology’ and Postmodern Pessimism.” In *Does Technology Drive History? The Dilemma of Technological Determinism*, edited by Merritt Roe Smith and Leo Marx. Cambridge, MA: MIT, 1994.
McClure, William B. “Technology and Command: Implications for Military Operations in the Twenty-first Century.” Occasional Paper No. 15. Maxwell AFB, AL: Center for Strategy and Technology, Air War College, Air University, July 2000.
McIntyre, Alison. “Doctrine of Double Effect.” In *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition),* edited by Edward N. Zalta. http://plato.stanford.edu/archives/fall2008/entries/double-effect (accessed21 February 2012).
Milgram, Stanley. “Some Conditions of Obedience and Disobedience to Authority.” *Human Relations* 18, no. 1 (February 1965): 57 – 76.
Oren, Nir, Simon Miles, and Michael Luck. “Representing Norms within Agent Systems. Paper presented. *5th Systems Engineering for Autonomous Systems Defense Technology Centre Technical Conference,* Edinburgh, UK, 14 – 15 July 2010: Paper B2.
Parasuraman, Raja. “Humans and Automation: Use, Misuse, Disuse, Abuse,” *Human Factors* 39, no. 2 (June 1997): 230 – 253.
Parasuraman, Raja, Thomas B. Sheridan, and Christopher D. Wickens. “A Model for Types and Levels of Human Interaction with Automation. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans* 30, no. 3 (May 2000): 286 – 297.
Reising, John. “Fitts’ Principles Still Applicable: Computer Monitoring of Fighter Aircraft Emergencies.” *Aviation Space and Environment Medicine* 53, no. 11 (Nov 1982): 1080 – 1084.
Rittel, Horst W.J. and Melvin M. Webber. “Dilemmas in a General theory of Planning.” *Policy Science* 4 (1973): 155 – 169.
Sargent, Robert G. “Verification and Validation of Simulation Models.” Paper Presented. *2005 IEEE Winter Simulation Conference*, Orlando, FL, 4 – 7 December 2005.
Schaffer, Ronald. “American Military Ethics in World War II: The Bombing of German Civilians.” *The Journal of American History* 67, no. 2 (September 1980): 318 – 334
Schultz, Colonel Timothy P. “Redefining Flight: How the Predecessors of the Modern United States Air Force Transformed the Relationship Between Airmen and Aircraft.” PhD diss., Duke University, 2007.

Schumann, Johann and Willem Visser. “Autonomy Software: V&V Challenges and Characteristics.” Paper presented. *2006 IEEE Aerospace Conference*, Big Sky, 4 – 11 March 2006.

Sharkey, Noel. “Cassandra or False Prophet of Doom: AI Robots and War.” *IEEE Intelligent Systems*, (July/August 2008): 14 – 17.

Sheridan, T.D. “Speculations on Future Relations Between Humans and Automation.” In *Automation and Human Performance*, edited by M. Mouloua. Mahwah, NJ: Lawrence Erlbaum Associates, 1996.

Singer, Peter W. “Tactical Generals: Leaders, Technology, and the Perils of Battlefield Micromanagement.” *Air & Space Power Journal* 23, no. 2 (Summer 2009). http://www.airpower.maxwell.af.mil/airchronicles/apj/apj09/sum09/singer.html (accessed 30 April 2012).

Singer, Peter W. “The Ethics of Killer Applications: Why Is It So Hard To Talk About Morality When It Comes to New Military Technology?” *Journal of Military Ethics* 9, no 4 (December 2010): 299 – 312.

Singer, Peter W. “Wired for War? Robots and Military Doctrine.” *Joint Forces Quarterly* no. 52 (1st Quarter 2009): 104 – 110. http://www.ndu.edu/press/lib/pdf/jfq-52/JFQ-52.pdf (accessed 30 April 2012).

Sinnott-Armstrong, Walter. “Consequentialism.” In *The Stanford Encyclopedia of Philosophy (Winter 2011 Edition)*, edited by Edward N. Zalta. http://plato.stanford.edu/archives/win2011/entries/consequentialism (accessed 20 February 2012).

Smith, Merritt Roe. “Technological Determinism in American Culture.” In *Does Technology Drive History? The Dilemma of Technological Determinism*, edited by Merritt Roe Smith and Leo Marx, 1 – 36. Cambridge, MA: MIT, 1994.

Sparrow, Robert. “Killer Robots.” *Journal of Applied Philosophy* 24, no. 1 (February 2007): 62 – 77.

Strawser, Bradley J. “Moral Predators: The Duty to Employ Uninhabited Aerial Vehicles.” *Journal of Military Ethics* 9, no. 4 (December 2010): 342 – 368.

Taylor, Glenn, Brian Stensrud, Susan Eitelman, Cory Dunham, and Echo Harger. “Toward Automating Airspace Management.” *IEEE Computational Intelligence for Security and Defense Applications* (April 2007): 124 – 130.

Thoms, Joanne. “Understanding the Impact of Machine Technologies on Human Team Cognition.” Paper presented. *4th Systems Engineering for Autonomous Systems Defense Technology Centre Technical Conference*, Edinburgh, UK, 7 – 8 July 2009. Paper B7.

Turing, Alan M. “Computing Machinery and Intelligence.” *Mind: A Quarterly Review of Psychology and Philosophy* 49, no. 236 (October 1950): 433 – 460.

Tvaryanas, Anthony P. “Human Systems Integration in Remotely Piloted Aircraft Operations.” *Aviation, Space, and Environmental Medicine* 77, no. 12 (December 2006), 1278 – 1282.
Winner, Langdon. “Do Artifacts Have Politics?” *Daedalus* 109, no. 1 (Winter 1980): 121 – 136.

Articles

Asimov, Isaac. “Runaround.” *Astounding Science Fiction*, March 1942, 94 – 103.
Boot, Max. “The Paradox of Military Technology.” *The New Atlantis*, Fall 2006. http://www.thenewatlantis.com/docLib/TNA14-Boot.pdf (accessed 30 April 2012).
Carlile, Col Christopher B. and Lt Col Gelnn Rizzi. “Robot Revolution: Revealing the Army’s UAS Road Map.” *Armed Forces Journal*, August 2010. http://www.armedforcesjournal.com/2010/08/4612789 (accessed 30 April 2012).
DARPA. “Urban Challenge.” http://archive.darpa.mil/grandchallenge/index.asp (accessed 26 March 2012).
Davis, Major Daniel L. “Who Decides: Man or Machine?” *Armed Forces Journal*, November 2007. http://www.armedforcesjournal.com/2007/11/3036753 (accessed 16 February 2012).
Deok-hyun, Kim. “Army Test Machine-Gun Sentry Robots in DMZ.” *Yonhap News Agency*, 13 July 2010. http://english.yonhapnews.co.kr/national/2010/07/13/14/0301000000AEN20100713007800315F.HTML# (accessed 26 March 2012).
Drew, Christopher. “Drones are Weapons of Choice in Fighting Qaeda.” *The New York Times*, 17 March 2009. http://www.nytimes.com/2009/03/17/business/17uav.html?pagewanted=all (accessed 30 March 2012).
Engelbrecht, Leon. “Did Software Kill Soldiers?” *IT Web*, 16 October 2007. http://www.itweb.co.za/index.php?option=com_content&view=article&id=6157&catid=96:defence-and-aerospace-technology (accessed 20 March 2012).
Finn, Peter. “A Future for Drones: Automated Killing.” *The Washington Post*, 19 September 2011. http://www.washingtonpost.com/national/national-security/a-future-for-drones-automated-killing/2011/09/15/glQAVy9mgK_story_1.html (accessed 19 March 2012).
ICRAC. “International Committee for Robot Arms Control.” www.icrac.co.uk (accessed 17 February 2012).

Independent Lens. “The Political Dr. Seuss.” http://www.pbs.org/independentlens/politicaldrseuss/seuss_fla.html (accessed 29 March 2012).

Joy, Bill. “Why the Future Doesn’t Need Us.” *Wired*, April 2000. http://www.wired.com/wired/archive/8.04/joy.html (accessed 16 February 2012).

Knight, Gavin. “March of the Terminators: Robot Warriors are no Longer Sci-Fi but Reality. So What Happens When They Turn Their Guns on Us?” *Daily Mail (UK)*, 15 May 2009. http://www.dailymail.co.uk/sciencetech/article-1182910/March-terminators-Robot-warriors-longer-sci-fi-reality-So-happens-turn-guns-us.html (accessed 20 March 2012).

Pachal, Peter. “IBM’s Watson Wins Jeopardy! Next Up: Fixing Health Care.” *PC Magazine*, 16 February 2011. http://www.pcmag.com/article2/0,2817,2380489,00.asp (accessed 26 March 2012).

Samsung Techwin. “SGR Series.” http://www.samsungtechwin.com/product/product_01_01.asp (accessed 26 March 2012).

Schmitt, Eric. “The World; The Powell Doctrine is Looking Pretty Good Again.” *The New York Times*, 4 April 1999. http://www.nytimes.com/1999/04/04/weekinreview/the-world-the-powell-doctrine-is-looking-pretty-good-again.html?pagewanted=all&src=pm (accessed 27 April 2012).

Shachtman, Noah. “Inside the Robo-Cannon Rampage.” *Wired Danger Room: What’s Next in National Security*, 19 October 2007. http://m.wired.com/dangerroom/2007/10/inside-the-robo/ (accessed 19 March 2012).

Shachtman, Noah. “Robot Cannon Kills 9, Wounds 14.” *Wired Danger Room: What’s Next in National Security*, 18 October 2007. http://m.wired.com/dangerroom/2007/10/robot-cannon-ki/ (accessed 19 March 2012).

Sharkey, Noel. “Grounds for Discrimination: Autonomous Robot Weapons.” *RUSI Defense Systems – Challenges of Autonomous Weapons*, October 2008: 86 – 89. http://www.rusi.org/downloads/assets/23sharkey.pdf (accessed 17 February 2012).

Simonite, Tom. “‘Robotic Rampage’ Unlikely Reason for Deaths.” *New Scientist*, 19 October 2007. http://www.newscientist.com/mobile/article/dn12812 (accessed 19 March 2012).

Singer, Peter W. “How the US Military Can Win the Robotic Revolution.” *Popular Mechanics*, 13 May 2010. http://www.popularmechanics.com/technology/military/robots/how-to-win-robot-military-revolution (accessed 30 April 2012).

Singer, Peter W. “In the Loop? Armed Robots and the Future of War.” *Defense Industry Daily*, 28 January 2009. http://www.defenseindustrydaily.com/In-the-Loop-Armed-Robots-and-the-Future-of-War-05267/ (accessed 30 April 2012).
Singer, Peter W. “We, Robot.” *Slate*, 19 May 2010. http://www.slate.com/id/2253692/ (accessed 30 April 2012).
TEXTRON Systems. “BLU-108 Submuntion.” http://www.textrondefense.com/assets/pdfs/datasheets/blu108_datasheet.pdf (accessed 15 February 2012).
Vanderbilt, Tom. “Let the Robot Drive: The Autonomous Car of the Future is Here.” *Wired*, 20 January 2012. http://www.wired.com/magazine/2012/01/ff_autonomouscars/all/1 (accessed 26 March 2012).
Weinberger, Casper. “Excerpts from Address of Weinberger.” *New York Times*, 29 November 1984.
“Were Lohatla Deaths an Accident?” *IOL News,* 25 January 2008. http://www.iol.co.za/news/south-africa/were-lohatla-deaths-an-accident-1.386941 (accessed 20 March 2012).
Wilson, George C. “Navy Missile Downs Iranian Jetliner.” *Washington Post*, 4 July 1988. http://www.washingtonpost.com/wp-srv/inatl/longterm/flight801/stories/july88crash.htm (accessed 22 March 2012).

Books

Air Force Operations & The Law: A Guide for Air, Space, and Cyber Forces. 2d ed. Maxwell AFB, AL: Air University Press, 2009. http://www.afjag.af.mil/shared/media/document/AFD-100510-059.pdf (accessed 30 April 2012).
Aquinas, St. Thomas. *The Summa Theologica, Part II, Question 40*. New York: Benziger Bros. edition, 1947. http://ethics.sandiego.edu/Books/Texts/Aquinas/JustWar.html (accessed 30 April 2012).
Aquinas, St. Thomas. *The Summa Theologica*. 2008 on-line edition, edited by Kevin Knight. http://www.newadvent.org/summa/3064.htm (accessed 21 February 2012).
Arkin, Ronald C. *Governing Lethal Behavior in Autonomous Robots*. New York: CRC Press, 2009.
Asimov, Isaac. *I, Robot*. Garden City, NY: Doubleday, 1950.
Bharucha-Reid, A.T. *Elements of the Theory of Markov Processes and Their Applications*. Mineola, NY: Dover Publications, 1960.
Boot, Max. *War Made New: Technology, Warfare, and the Course of History 1500 to Today*. New York: Gotham, 2006.

Bousquet, Antione J. *The Scientific Way of Warfare: Order and Chaos on the Battlefields of Modernity*. New York, NY: Columbia University, 2009.
Brigety, Reuben E., II. *Ethics, Technology, and the American Way of War: Cruise Missiles and US Security Policy*. New York: Routledge, 2007.
Carroll, Lewis. *Alice in Wonderland*. Scituate, MA: Digital Scanning, 2007.
Clausewitz, Carl von. *On War*. Edited and translated by Michael Howard and Peter Paret. Princeton, NJ: Princeton University, 1984.
Craig, Campbell. *Destroying the Village: Eisenhower and Thermonuclear War*. New York: Columbia University, 1998.
Creveld, Martin van. *Command in War*. Cambridge, MA: Harvard University, 1985.
Davis, Tami. *Rhetoric and Reality in Air Warfare: The Evolution of British and American Ideas About Strategic Bombing, 1914 – 1945*. Princeton, NJ: Princeton University, 2002.
Döner, Dietrich. *The Logic of Failure–Why Things Go Wrong and What We Can Do to Make Them Right*. Translated by Rita and Robert Kimber. New York: Metropolitan Books, 1996.
Douhet, Giulio. *The Command of the Air*. Edited and translated by Joseph Harahan and Richard Kohn. Tuscaloosa, AL: University of Alabama, 2009.
Dower, John W. *War without Mercy: Race and Power in the Pacific War*. New York: Pantheon, 1986.
Fallows, James. *National Defense*. New York: Random House, 1981.
Fox, John and Subrata Das. *Safe and Sound: Artificial Intelligence in Hazardous Applications*. Cambridge, MA: MIT Press, 2000.
Franks, Tommy. *American Solider*. New York: HarperCollins, 2004.
George, Alexander L. *Forceful Persuasion: Coercive Diplomacy as an Alternative to War*. Washington, DC: United States Institute of Peace, 1994.
George, Alexander L. and William E. Simons, eds. *The Limits of Coercive Diplomacy*. 2d ed. Boulder, CO: Westview, 1994
Gladwell, Malcolm. *Blink: The Power of Thinking Without Thinking*. New York: Little, Brown and Company, 2005.
Gray, Colin S. *Explorations in Strategy*. Westport, CT: Praeger, 1996.
Gray, Colin S. *Modern Strategy*. New York: Oxford University, 1999.
Grosman, Dave. *On Killing: The Psychological Cost of Learning to Kill in War and Society*. New York: Back Bay Books, 2009.
Hambling, David. *Weapons Grade: How Modern Warfare Gave Birth to Our High-Tech World*. New York: Carroll & Graff, 2005.
Holley, Irving B., Jr. *Ideas and Weapons*. Washington, DC: Government Printing Office, 1997.
Ignatieff, Michael. *Virtual War: Kosovo and Beyond*. New York: Metropolitan Books, 2000.

Jones, Seth G. *In the Graveyard of Empires: America's War in Afghanistan*. New York: Norton, 2010.
Kalyvas, Stathis N. *The Logic of Violence in Civil War*. New York: Cambridge University, 2006.
Kant, Immanuel. *Grounding for the Metaphysics of Morals*. Edited and translated by James W. Ellington. Indianapolis, IN: Hackett Publishing, 1993.
Kometer, Michael W. *Command in War: Centralized Versus Decentralized Control of Combat Airpower*. Maxwell AFB, AL: Air University, 2007.
Krakauer, John. *Where Men Win Glory: The Odyssey of Pat Tillman*. New York: Doubleday, 2009.
Kranz, Gene. *Failure is not an Option: Mission Control from Mercury to Apollo 13 and Beyond*. New York: Simon & Schuster, 2000.
Krishnan, Armin. *Killer Robots: Legality and Ethicality of Autonomous Weapons*. Burlington, VT: Ashgate, 2009.
Kurzweil, Ray. *The Singularity is Near: When Humans Transcend Biology*. New York: Viking, 2005.
Lonsdale, David J. *The Nature of War in the Information Age: Clausewitzian Future*. New York, NY: Frank Cass, 2004.
Luttwak, Edward. *Strategy: The Logic of War and Peace*. Cambridge, MA: Harvard University, 2001.
Mansfield, Edward and Jack Snyder. *Electing to Fight: Why Emerging Democracies go to War*. Cambridge, MA: MIT, 2005.
Mao Tse-Tung. *Selected Works of Mao Tse-Tung*. Vol. 2. Peking, China: Foreign Language Press, 1965; new imprint Digital Reprints 2007.
McNeill, William H. *The Pursuit of Power: Technology, Armed Force, and Society since A.D. 1000*. Chicago, IL: University of Chicago, 1982.
Mindell, David A. *Between Human and Machine: Feedback, Control, and Computing before Cybernetics*. Baltimore, MD: Johns Hopkins University, 2002.
Moltke, Helmuth Graf von. *Moltke on the Art of War: Selected Writings*. Translated by Daniel Hughes. New York: Ballantine, 1993.
Moore, Mike. *Twilight War: The Folly of U.S. Space Dominance*. Oakland, CA: The Independent Institute, 2008.
Nye, Joseph. *The Future of Power*. New York: Public Affairs, 2011.
Osinga, Frans. *Science, Strategy and War: The Strategic Theory of John Boyd*. New York: Routledge, 2007.
Pape, Robert. *Bombing to Win: Air Power and Coercion in War*. Ithaca, NY: Cornell University, 1996.
Pierce, Watson O'D. *Air War: Its Psychological, Technical, and Social Implications*. New York: Modern Age Books, 1939.
Powell, Colin L. *My American Journey*. New York: Random House, 1995.
Randolph, Stephen. *Powerful and Brutal Weapons*. Cambridge, MA: Harvard University, 2007.

Rosen, Stephen P. *Winning the Next War: Innovation and the Modern Military*. Ithaca, NY: Cornel University, 1991.
Sagan, Scott D. *The Limits of Safety: Organizations, Accidents, and Nuclear Weapons*. Princeton, NJ: Princeton University, 1993.
Schelling, Thomas C. *Arms and Influence*. 2008 ed. New Haven, CT: Yale University, 2008.
Shakespeare, William. *The Tempest.* Edited by Virginia M. Vaughan and Alden T. Vaughan. London: The Arden Shakespeare, 1999.
Shelley, Mary. *Frankenstein*. New York: SoHo Books, 2010.
Sherry, Michael. *The Rise of American Air Power: The Creation of Armageddon*. New Haven, CT: Yale University, 1987.
Singer, Peter W. *Wired for War: The Robotics Revolution and Conflict in the 21st Century*. New York: Penguin, 2009.
Smyth, Henry D. *Atomic Energy for Military Purposes.* York, PA: Maple Press, 1945.
Sun Tzu. *The Illustrated Art of War*. Translated by Samuel Griffith. New York: Oxford University, 2005.
Verne, Jules. *From the Earth to the Moon: Direct in Ninety-Seven Hours and Twenty Minutes: and a Trip Round It*. New York: Charles Scribner's Sons, 1890.
Waldrop, M. Mitchell. *Complexity: The Emerging Science at the Edge of Order and Chaos*. New York: Simon & Schuster, 1992.
Wallach, Wendell and Colin Allen. *Moral Machines: Teaching Robots Right from Wrong*. New York: Oxford University Press, 2009.
Walzer, Michael. *Just and Unjust Wars: A Moral Argument with Historical Illustrations*. New York: Basic Books, 1977.
Weigley, Russel F. *The American Way of War: A History of United States Military Strategy and Policy*. Bloomington, IN: Indiana University, 1973.
Wells, H.G. *The War in the Air*. New York: MacMillan, 1907.
Wells, H.G. *War of the Worlds*. New York: SoHo Books, 2010.
Werrell, Kenneth P. *Chasing the Silver Bullet: U.S. Air Force Weapons Development from Vietnam to Desert Storm*. Washington, DC: Smithsonian Books, 2003.
White, Lynn, Jr. *Medieval Technology and Social Change*. New York: Oxford, 1964.
Wiener, Norbert. *Cybernetics, or Control and Communication in the Animal and the Machine*. 2d ed. New York: MIT Press, 1961.
Winner, Langdon. *Autonomous Technology Technics-out-of-Control as a Theme in Political Thought*. Boston, MA: MIT, 1978.
Yarger, Harry R. *Strategy and the National Security Professional: Strategic Thinking and Strategy Formulation in the 21st Century*. Westport, CT: Praeger, 2008.

Briefings/Speeches

Bradley, General Omar N. “An Armistice Day Address.” Address. Chamber of Commerce, Boston, MA, 10 November 1948. Text available at http://www.opinionbug.com/2109/armistice-day-1948-address-general-omar-n-bradley (accessed 21 March 2012

Department of Defense. “Background Briefing on Air-Sea Battle.” 9 November 2011. www.defense.gov/transcript.aspx?transcriptid=4932 (accessed 16 February 2012).

Tether, Tony. Director, Defense Advanced Research Projects Agency. Statement before the Subcommittee on Military Research and Development, Committee on Armed Services, Washington, DC, 26 June 2001. Text available at www.darpa.mil/WorkArea/DownloadAsset.aspx?id=1781 (accessed 3 May 2012).

Films

James Cameron. *The Terminator*. Los Angeles, CA: MGM, 1984.

James Cameron. *Terminator 2: Judgment Day*. Los Angeles, CA: Carolco, 1991.

Jonathan Mostow. *Terminator 3: Rise of the Machines*. Los Angeles, CA: Warner Brothers, 2003.

Joseph McGinty Nichol. *Terminator Salvation*. Los Angeles, CA, Warner Brothers, 2009.

Government Documents

ACTS. *Air Force* text 1931. In AFHRC, decimal file no. 248.101-16.

ACTS. “Air Offensive Characteristics” *Air Force Air Warfare*, 1 Feb 1938. In AFHRC, decimal file 248.101-1.

ACTS. “Character and Strategy of Air Power.” *Air Force: Part One*, 1 Dec 1935. In AFHRC, decimal file no. 248.101-1.

ACTS. “International Aerial Regulations” text, 1933-34. In AFHRC, decimal file no. 248.101-16.

Foulois, Major General Benjamin D., USAF (Ret). “Early Flying Experience in Army Airplane No. 1 (1909-1910-1911).” Unpublished article, 1960. Manuscript series 17, box 6 Clark Special Collections Branch, USAF Academy Library, CO.

Handbook of Instructions for Airplane Designers. 7th ed., vol. 1. Wright Field, OH: U.S. Army Air Corps, 1934. History Office, Air Force Material Command (AFMC), Wright Patterson Air Force Base, Ohio

Ministry of Defense. Joint Doctrine Note 2/11 *The UK Approach to Unmanned Aircraft Systems*. London: Office of the Assistant Head Air and Space (Development, Concepts and Doctrine), 30 March 2011.

"Pilot Training Manual for the B-17 Flying Fortress." HQ Army Air Forces, Office of Flying Safety. Special manuscript series 603, box 3, Clark Special Collections Branch, USAF Academy Library, CO.
Reilly, Jeffrey M. *Design: Distilling Clarity for Decisive Action*. Maxwell Air Force Base, AL: Air Command and Staff College [Department of Joint Warfare Studies], October 2011.
USAF Test Pilot School. "Technology and Automation." *Systems Phase Text Book Chapter 3 – Human Factors*. Edwards AFB, CA: Air Force Material Command, July 2002.

Legal Documents

Convention for the Amelioration of Condition of the Wounded and Sick in Armed Forces in the Field, 12 August 1949. 75 U.N.T.S. 970.
Convention for the Amelioration of the Condition of the Wounded, Sick, and Shipwrecked Members of the Armed Forces at Sea, 12 August 1949. 75 U.N.T.S. 971
Convention Relative to the Treatment of Prisoners of War, 12 August 1949. 75 U.N.T.S. 972.
Convention Relative to the Protection of Civilian Persons in Time of War, 12 August 1949. 75 U.N.T.S. 973.
Hague Convention (IV) respecting the Laws and Customs of War on Land and its Annex: Regulations Concerning the Laws and Customs of War on Land. The Hague, 18 October 1907. http://www.icrc.org/ihl.nsf/FULL/195?OpenDocument (accessed 23 January 2012).
Instructions for the Government of Armies of the United States in the Field (Lieber Code), 24 April 1863. http://www.icrc.org/ihl.nsf/FULL/110?OpenDocument (accessed 23 January 2012).
International Court of Justice. *Legality of the Threat or Use of Nuclear Weapons*. Advisory Opinion, 8 July 1996, http://www.icj-cij.org/docket/index.php?p1=3&p2=4&k=e1&p3=4&case=95 (accessed 30 April 2012).
Jacobellis v. Ohio. 378 U.S. 184, 1964.
Prosecutor v. Tadic. Case No. IT-94-1-I. Decision on the Defense Motion for Interlocatory Appeal on Jurisdiction, 2 October 1995. http://www.icty.org/x/cases/tadic/acdec/en/51002.htm (accessed 10 January 2012).
Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts (Protocol I), 8 June 1977. 1125 U.N.T.S. 25. http://treaties.un.org/doc/Publication/UNTS/Volume%201125/v1125.pdf (accessed 26 January 2012).

Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I), 8 June 1977. 1125 U.N.T.S. 26.
Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I), 8 June 1977. 1125 U.N.T.S. 29.
United Nations. *Charter of the United Nations*. http://www.un.org/en/documents/charter/chapter1.shtml (accessed 10 January 2012).
UN General Assembly Resolution 2444, 18 December 1968. http://www.icrc.org/ihl.nsf/FULL/440?OpenDocument (accessed 23 January 2012).

Reports

Arkin, Ronald. *Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture*. Technical Report GIT-GVU-07-11. Atlanta, GA: Georgia Institute of Technology, 2008.
Canadian Expeditionary Forces. *Board of Inquiry Minutes of Proceedings A-10A Friendly Fire Incident 4 September 2006, Panjwayi District, Afghanistan.* Ottowa, Canada: Department of National Defense, 13 July 2007. http://www.forces.gc.ca/site/focus/opmedusa/A10_BOI_Report_e.pdf (accessed 26 April 2012).
Canning, Paul S. *A Definitive Work on Factors Impacting the Arming of Unmanned Vehicles*. NSWCDD TR-05/36. Dahlgren, VA: Department of the Navy, May 2005.
Chappelle, Wayne, Kent McDonald, and Raymond King. *Psychological Attributes Critical to the Performance of MQ-1 Predator and MQ-9 Reaper US Air Force Sensor Operators*. AFRL-SA-BR-TR-2010-0007. Brooks City-Base, TX: Air Force Research Laboratory, 2010.
Committee on Automation in Combat Aircraft. *Automation in Combat Aircraft*. Washington, DC: National Research Council, 1982.
Dahm, Werner J.A. *Technology Horizons: A Vision for Air Force Science and Technology During 2010-2030*. AF/ST-TR-10-01-PR. Washington, DC: Department of the Air Force, 15 May 2010.
Department of the Air Force. *United States Air Force Unmanned Aircraft Systems Flight Plan 2009 – 2047*. Washington, DC: Headquarters, United States Air Force, 18 May 2009.
Department of the Army. *Mental Health Advisory Team IV Operation Iraqi Freedom 05 – 07 Final Report*. Washington, DC: Office of the Surgeon General, 17 November 2006. http://www.armymedicine.army.mil/reports/mhat/mhat_iv/MHAT_IV_Report_17NOV06.pdf (accessed 30 April 2012).

Department of Defense. *FY2009 – 2034 Unmanned Systems Integrated Roadmap*. Washington, DC: Office of the Under Secretary of Defense (Acquisition, Technology and Logistics), 2 April 2009.

Department of Defense. *Patriot System Performance Report Summary*, Defense Science Board Task Force Report. Washington, DC: Department of Defense, January 2005. http://www.acq.osd.mil/dsb/reports/ADA435837.pdf (accessed 3 May 2012).

Department of Defense. *Quadrennial Defense Review Report*. Washington, DC, February 2010.

Department of Defense. *Unmanned Systems Safety Guide for DOD Acquisition*. Washington, DC: Office of the Undersecretary of Defense (Acquisition, Technology, & Logistics), 27 June 2007.

Fitts, Paul M. *Human Engineering for an Effective Air-Navigation and Traffic-Control System*. Washington, DC: National Research Council, March 1951.

Government Accounting Office. *Patriot Missile Software Problem*. GAO/IMTEC-92-26. Washington, DC: General Accounting Office, 1992.

Huang, Hui-Min, Elena Messina, and James Albus. *Autonomy Levels for Unmanned Systems (ALFUS) Framework Volume II: Framework Models Version 1.0*. NIST Special Publication 1011-II-1.0. Gaithersburg, MD: National Institute of Standards and Technology, December 2007.

Karman, Theodore von. *Toward New Horizons: Science, Key to Air Supremacy*. Wright Field, Dayton, OH: Headquarters Air Material Command, May 1945. Document is now declassified.

Lin, Patrick, George Bekey, and Keith Abney. *Autonomous Military Robotics: Risk, Ethics, and Design*. Report for the Office of Naval Research. San Luis Obispo, CA: California Polytechnic State University, 20 December 2008.

Ministry of Defense. *Aircraft Accident to Royal Air Force Tornado GR MK4A ZG710*. London, UK: Directorate of Air Staff, March 2004.

Pickering, W.H. "Automatic Control of Flight." In *Guided Missiles and Pilotless Aircraft: A Report of the AAF Scientific Advisory Group*, edited by H.L. Dryden, W.H. Pickering, H.S. Tsien, and G.B. Schubauer. Wright Field, Dayton, OH: Headquarters Air Material Command, May 1946. Document is now declassified

United Nations General Assembly. *Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, Philip Alston*, Addendum, *Study on Targeted Killings*. A/HRC/14/24/Add.6, 28 May 2010.

United States Central Command. *Investigation of Suspected Friendly Fire Incident Near An Nasiriayah, Iraq, 23 March 2003*. Tampa, FL: US Central Command, 6 March 2004. The document is now declassified.

U.S.-Canada Power System Outage Task Force. *Final Report on the August 14, 2003 Blackout in the United States and Canada: Causes and Recommendations*. Washington, DC: Department of Energy, April 2004.